



**Universitat de Lleida**

Document downloaded from:

<http://hdl.handle.net/10459.1/69771>

The final publication is available at:

<https://doi.org/10.1021/acs.jpca.7b05089>

Copyright

(c) American Chemical Society, 2017

# Ionization and Conformational Equilibria of Citric Acid: Delocalized Proton Binding in Solution

Sergio Madurga,<sup>\*,†</sup> Miroslava Nedyalkova,<sup>‡</sup> Francesc Mas,<sup>†</sup> and Josep Lluís Garcés<sup>¶</sup>

<sup>†</sup>*Materials Science and Physical Chemistry Department & Research Institute of Theoretical and Computational Chemistry (IQTCUB) of Barcelona University (UB), C/ Martí i Franquès, 1. 08028 Barcelona (Catalonia, Spain)*

<sup>‡</sup>*Inorganic Chemistry Department, Faculty of Chemistry and Pharmacy "St Kliment Ohridski", University of Sofia, 1 James Bourchier Blvd. 1164, Sofia (Bulgaria)*

<sup>¶</sup>*Chemistry Department and AGROTECNIO, University of Lleida (UdL), Rovira Roure, 191, 25198, Lleida (Catalonia, Spain)*

E-mail: s.madurga@ub.edu

## Abstract

The micro-speciation of citric acid is studied by analyzing NMR titration data. When the site binding (SB) model, which assumes fully localized proton binding to the carboxylic groups, is used to obtain microscopic energy parameters (dissociation constants, pair and triplet interaction energies between charged carboxylate groups), contradictory results are obtained. The resulting macroscopic constants are in very good agreement with the values reported in the literature using potentiometry. However, the found pair interaction energy between the terminal carboxylates and the triplet interaction energy are physically meaningless. In order to solve this apparent contradiction we consider the possibility of delocalized

proton binding, so that the proton can be exchanged at high velocity in the NMR time scale through short, strong, low-barrier (SSLB) hydrogen bonds. With this aim, *ab initio* MP2 calculations using the SMD polarizable continuum model for the solvent were performed and the fully roto-microspeciation elucidated. Firstly, fully localized proton binding was assumed, and the resulting micro-state probabilities are in reasonable agreement with those reported in previous works which use selective blocking of the carboxylic groups. They are, however, in clear disagreement with the microstate probabilities derived from the NMR titration data, which predict, within a very narrow confidence interval, a unique micro-species for the symmetric di-ionized form. Moreover, counterintuitively, the interaction between terminal charged groups is much larger than that between central and terminal groups. As a consequence, we have explored the possibility of delocalized proton binding by calculating the energy of intermediate proton positions between two carboxylic groups. The results reveal that the exchange of the proton through the hydrogen bonds is in some cases produced without energetic barrier. This effect is specially relevant in the di-ionized form, with all the most stable conformations forming a SSLB, which together would constitute the only micro-state detected by NMR. An alternative reaction scheme for the ionization process, based on proton delocalization, is proposed.

## Introduction

Citric acid is an important molecule involved in a broad range of biological, industrial and environmental processes.<sup>1</sup> For instance, it is a natural preservative and flavoring agent used in the food industry<sup>2</sup> or a removable agent for heavy metal ions from waste-water.<sup>3</sup> Citrate is also relevant in colloidal sciences, where it is used as a common stabilizer and a reducing agent in the synthesis of gold nanoparticles, or as a coating on nanomaterials for biological applications.<sup>4</sup> It is also involved in differential adsorption on crystal surfaces leading to differential growing of crystal geometries.<sup>5</sup> All these properties are pH-dependent, so that a detailed knowledge of the ionization mechanism of this small molecule is necessary. However, most of the studies on

the ionization properties of citric acid have been approached within the macroscopic picture. Each protonation step is described by an equilibrium reaction characterized by a macroscopic equilibrium constant. Citric acid have three ionizable carboxylic groups, and the corresponding macroscopic dissociation constants have been mostly measured using potentiometry, in a wide range of ionic strengths and temperatures, in presence of different ground electrolytes.<sup>6-9</sup> However, in the macroscopic approach, no information is provided about the population of the micro-states, for which the protonation state of every individual site is specified.<sup>10-12</sup> In the case of citric acid, there is still controversy about the micro-state populations compatible with the macroscopic dissociation constants.<sup>13-15</sup>

Micro-speciation of ionizable molecules, and in particular that of citric acid, has mainly been performed within the context of the site-binding (SB) model, for which every protonating site can adopt two possible states, *i.e.*, protonated or deprotonated. This means that the proton is localized and ascribed to a unique protonating group, *i.e.*, it can not be shared by two ionizable sites. The SB model has been successfully used in a wide range of molecules, ranging from small molecules to polyelectrolytes and proteins, both for proton and metal binding.<sup>11,16,17</sup> Within this framework, two approaches have been applied to the determination of the molecular micro-speciation. In the so-called 'deductive' methods, the molecule is modified so that one or several ionizable groups are blocked. The non-blocked groups are supposed to have the same ionization properties as in the original molecule. This assumption is the weakest part of the approach since changes in the chemical structure of the molecule could modify the conformational and thus the chemical environment of the rest of the groups,<sup>12,18</sup> including the possible alteration or even destruction of hydrogen bonds. In the case of citric acid, several authors have selectively blocked the carboxylic groups forming the corresponding methyl-esters. The reported results, although consistent with the value of the macroscopic constants, differ in the micro-state populations.<sup>13-15</sup>

A second approach applied to the determination of the micro-speciation of small molecules is the use of spectroscopic techniques which are sensitive to the protonation of specific sites and do not perturb the conformational structure of the molecule. Among them, high resolution

1  
2  
3  
4 NMR is probably the most versatile and has been widely used in the micro-speciation of a vast  
5 array of ionizable molecules ranging from proteins to small molecules.<sup>10–12,19–22</sup> The chemical  
6 shift of a nucleus located in the proximity of an ionizable group is sensitive to its protonation  
7 state. Actually, it has been shown that, if the conformational equilibria are fast enough, the  
8 chemical shift is a linear combination of the degree of protonation of the neighbouring sites.<sup>10,23</sup>  
9  
10 As a consequence, measures of the chemical shift vs pH provide information about the micro-  
11 speciation of the molecule. Citric acid is an AB spin system with two sets of coupled methylene  
12 protons (Fig. 1). Moore and Sillerud measured the pH dependence of the high resolution NMR  
13 spectra of citric acid, including both the chemical shift and the spin-spin coupling,<sup>24</sup> although no  
14 micro-speciation analysis was reported.

25  
26 However, situations have been reported for which the proton can be hardly ascribed to an in-  
27 dividual donor group, the so called short, strong, low-barrier hydrogen bonds (SSLB-HB). SSLB-  
28 HB present a very low energetic barrier between the asymmetric forms  $A - O - H \cdots O - B$   
29 and  $A - O \cdots H - O - B$ , so that the location of the proton should be regarded as indetermi-  
30 nate. The resulting resonance structure  $A - O \cdots H \cdots O - B$  seems to show extra strength or  
31 stability, and has attracted great interest in the last years because of their possible role in the sta-  
32 bilization of transition states in enzymatic reactions.<sup>25</sup> The existence of SSLB-HB has been widely  
33 accepted in solid state. For instance, neutron diffraction analysis of the crystals of monoanions of  
34 dicarboxylic acids, such as maleate or phthalate, show that the proton is centered between the two  
35 donor oxygens.<sup>26–28</sup> The same conclusion has also been achieved in X-ray crystal analysis, where  
36 the separations between the two donor oxygen atoms are characteristically short, compared to  
37 the usual, asymmetric, hydrogen bonds.<sup>29,30</sup> In solution, however, the matter is more controver-  
38 sial. The study of UV and IR spectra of poly(fumaric) and poly(maleic) acids by Kawaguchi *et*  
39 *al.* suggests the existence of shared protons between consecutive carboxylic groups, forming a  
40 kind of third species besides the “fully” protonated and “fully” deprotonated groups.<sup>31</sup> Recent  
41 research has shown the existence of dicoordinated proton in condensed media. An archetypal  
42 example would be the bisdiethyletherate salts, composed by  $H(OEt_2)_2^+$  cations and weakly coor-  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

dinating anions, being formed when strong acids are dissolved in diethyl ether.<sup>32</sup> IR spectroscopy has shown that protons in liquid water form dicoordinated or even more complex structures of the kind of  $[\text{H}(\text{H}_2\text{O})_n]^+$ . Conversely, studies based on the NMR method of isotopic perturbation report that some of the dicoordinated structures existent in solid state, such as maleate, phthalate, among others, are, in solution, two tautomers in equilibrium differing in the solvation structure (or solvatomers) of the carboxylic groups, rather than a single symmetric species. However, even in this case, the H-bond of the solvatomers is not necessarily described by a double-well potential, but with a single-well potential with an additional contribution because of the instantaneous solvation.<sup>28,33</sup> Thus, certain degree of bond asymmetry seems to be compatible up to some extent with proton delocalization. QM/MM calculations with explicit solvent have also shown the existence of a very low energetic barrier in the H-bond of the hydrogen-phthalate anion.<sup>34</sup>

Regarding citric acid, previous *ab initio* calculations were performed to study its conformational properties without the consideration of delocalized hydrogen bonds. Khan *et al.*<sup>35</sup> generated conformations of citric acid and its dissociated forms using a Monte Carlo method with a classical force-field. The low energy conformers were optimized with the B3LYP/6-31+G level in the gas phase and the polarized continuum model was used to perform single point solvent calculations. Wrigth *et al.*<sup>36</sup> performed Carr-Parrinello molecular dynamics simulations for the completely deprotonated form of citric acid to obtain force-field parameters.

In the present work, we consider the possibility of delocalized proton in some of the ionization micro-states of citric acid by means of NMR titration data analysis, on the one side, and *ab initio* calculations using the SMD polarizable continuum model for the solvent,<sup>37</sup> on the other. A brief summary of both methodologies is provided in section 2. The available NMR titration data reported by Moore and Sillerud in ref.<sup>24</sup> are analyzed in section 3 within the SB model, *i.e.*, a bound proton is ascribed to a unique carboxylic group. Microscopic ionization constants, interactions between ionized groups and micro-state probabilities are obtained. The robustness of the best fitted parameters is tested by using Monte Carlo simulations of synthetic data sets. The results, however, seem to be in part inconsistent with fully localized proton binding. Some of the obtained

interaction energies, within a very narrow confidence interval, show nonphysical values, despite the macroscopic constants are in very good agreement with those previously obtained by using potentiometric titration under similar conditions. In particular, counterintuitively, the interaction between terminal charged groups is much larger than that between central and terminal groups. This fact could be explained because the reaction scheme does not take into account the possibility of delocalized proton binding in some roto-microstates, so that artificial effective interaction energies are found. This possibility is discussed in section 4, where the *ab initio* calculations are presented. With this aim, we perform an exhaustive scan of the most probable conformational states within a particular macrostate (or roto-microstates), with special regard to those involving H-bonds. The computations suggest that delocalized proton binding could be taking place in the mono-ionized and di-ionized roto-microstates, with the proton moving without energetic barrier between the ionized carboxylic group. This phenomenon could explain the anomalies detected in the NMR analysis.

## Methodologies

### Macrostates, Microstates and Roto-microstates. Site Binding (SB) model

Let us firstly assume fully localized proton binding, so that a bound proton is ascribed to a unique ionizable group. This is the starting hypothesis of the so-called site binding (SB) model. A particular conformational and ionization state, or *roto-microstate*, of a molecule with  $n$  protonated sites is fully characterized by the probability  $p_n(c, q, a_H)$ , where  $a_H$  denotes the proton activity,  $q$  the ionization state and  $c$  a particular conformational state. The ionization state  $q$  is defined by a set of variables  $q = \{q_1, q_2, \dots, q_N\}$ , where  $N$  is the total number of ionizable sites.  $q_i$  can take two possible values:  $q_i = 1$  is the site is charged (*i.e.* deprotonated) and  $q_i = 0$  otherwise.  $n = N - \sum_{i=1, \dots, N} q_i$  is the total number of protonated groups and define a *macrostate*  $n$ .

It can be shown that  $p_n(c, q, a_H)$  can be expressed in the form<sup>38</sup>

$$p_n(c, q, a_H) = \rho_n(c, q) P_n(a_H) \quad (1)$$

$\rho_n(c, s)$  represents the conditional probability of a roto-microstate provided that belongs to a macrostate  $n$ .  $P_n(a_H)$  is the macrostate probability and it is given by

$$P_n(a_H) = \frac{\overline{K}_n a_H^n}{\Xi} \quad (2)$$

where  $\overline{K}_n$  are the macroscopic protonation constants and  $\Xi = \sum_{i=0, \dots, N} \overline{K}_i a_H^i$  (with  $\overline{K}_0 = 1$ ) is the semi-grand canonical partition function.  $\overline{K}_n$  corresponds to the process of binding  $n$  protons to the fully deprotonated molecule. The set of  $\overline{K}_n$  characterizes the macroscopic description, which does not differentiate between individual protonated groups. Note that, according to eqn. (1) the pH-dependence of the roto-microstate probabilities comes only from the macrostates probabilities, while the conditional probabilities are pH independent.<sup>11,38</sup> For carboxylic acids as citric acid, the experimental quantities currently reported are the dissociation  $pK_i$ -values.  $pK_i$  corresponds to the process of releasing the  $i$ th proton, i.e., the deprotonation of the species with  $N-i+1$  protons.  $K_i$  and  $\overline{K}_n$  are related by  $pK_i = -\log K_i = \log(\overline{K}_{N-i+1}/\overline{K}_{N-i})$ . Finally, a microstate  $s$  is defined as the set of roto-microstates with the same ionization state.<sup>38</sup> The microstate probability is given by

$$\pi_n(q) = \sum_c \rho_n(c, q) \quad (3)$$

and represents the conditional probability of a microstate  $s$  within a macrostate  $n$ . The microstate probabilities are related to the microstate free energy as

$$\pi_n(q) = \frac{e^{-\beta F(q)}}{\sum_s e^{-\beta F(q)}} \quad (4)$$



which can be expressed in the form of the so-called cluster expansion

$$\frac{\beta F(q)}{\ln 10} = \sum_i p k_i q_i + \sum_{j>i} \epsilon_{ij} q_i q_j + \sum_{k>j>i} \lambda_{ijk} q_i q_j q_k + \dots \quad (5)$$

where  $\beta = 1/k_B T$ ,  $p k_i$  is the microscopic dissociation constant of the site  $i$ ;  $\epsilon_{ij}$  is the pair interaction energy between the charged sites  $i$  and  $j$ ;  $\lambda_{ijk}$  represents the triplet interaction energy, etc.  $p k_i$ ,  $\epsilon_{ij}$ ,  $\lambda_{ijk}$ , the cluster parameters, do not depend on the conformational state, and they are proper averages of the roto-microstates.<sup>38,39</sup> Note that the choice of the protonation variables is not unique. For instance, for polycations such as polyamines, a better choice is the set of variables  $s_i = 1 - q_i$ , so that the interaction terms are non zero when the sites are charged, *i.e.* protonated. For polyanions, however, the choice of the variables  $q_i$  seems more natural since the carboxylic groups are charged when they are deprotonated. In any case, the choice of the protonation variables does not affect the resulting thermal averages.<sup>11</sup> For citric acid, eqn. 5 can be simplified using the symmetry of the molecule. Since the two terminal carboxylic groups are equivalent, the cluster expansion remains

$$\frac{\beta F(q)}{\ln 10} = p k_1 (q_1 + q_3) + p k_2 q_2 + \epsilon_{12} (q_1 q_2 + q_2 q_3) + \epsilon_{13} q_1 q_3 + \lambda_{123} q_1 q_2 q_3 \quad (6)$$

so that five parameters are necessary to fully characterize the micro-speciation of citric acid, as depicted in Fig. 2a. They can be experimentally obtained by the techniques outlined in the introduction. In this work the pH dependence of the NMR chemical shifts of citric acid is analyzed.

## NMR titrations and SB model

If the conformational and protonation equilibria are fast enough, it can be rigorously shown that the chemical shift of a particular atom  $\alpha$  close to a ionizable group is given by the average of the chemical shifts of the possible roto-microstates of the molecule.<sup>40</sup> The simple form of the NMR spectrum of citric acid, shown in Fig. 1 indicates that this hypothesis is fulfilled by citric acid.<sup>24</sup>

The chemical shift of a  $^1\text{H}$  nucleus close to an ionizable group at a given pH,  $\delta_\alpha(a_{\text{H}})$ , is then given by<sup>12</sup>

$$\delta_\alpha(a_{\text{H}}) = \sum_{c,q} \delta'_\alpha(c,q) p_n(c,q,a_{\text{H}}) \quad (7)$$

where  $\delta'_\alpha(c,q)$  is the chemical shift corresponding to the atom  $\alpha$  when the molecule is in a particular roto-microstate  $(c,q)$ . Eqn. (7) together with (1) leads to

$$\delta_\alpha(a_{\text{H}}) = \sum_{n=0,\dots,N} \tilde{\delta}_{\alpha n} P_n(a_{\text{H}}) \quad (8)$$

where  $\tilde{\delta}_{\alpha n} = \sum_{c,s} \delta'_\alpha(c,q) \rho_n(c,q) \delta_{n,\sum_{i=1,\dots,N} s_i}$ ,  $s_i = 1 - q_i$  and  $\delta_{n,\sum_{i=1,\dots,N} s_i}$  is the Kronecker delta, which takes the value 1 when there are exactly  $n$  protons, and 0 otherwise. Note that  $\tilde{\delta}_{\alpha n}$  corresponds to the average of the chemical shifts of the roto-microstates corresponding to the same macrostate. Eqn. (8), together with (2), allows determining the macroscopic constants by fitting the experimental chemical shifts as a function of the pH. This information can be also available by potentiometric titration. However, in order to obtain microscopic information, it is necessary to express  $\delta_\alpha(a_{\text{H}})$  as a function of the site-specific degrees of protonation,  $\theta_i(a_{\text{H}})$ ,  $i = 1, \dots, N$ . They are related to the macroscopic probabilities by

$$\theta_i(a_{\text{H}}) = \sum_{n=1,\dots,N} A_{in} P_n(a_{\text{H}}) \quad (9)$$

where the coefficients  $A_{in}$  are given by<sup>10</sup>

$$A_{in} = \sum_s s_i \pi_n(q) \delta_{n,\sum_{i=1,\dots,N} s_i} \quad (10)$$

Expressing the macroscopic probabilities in terms of  $\theta_i(a_{\text{H}})$  by inverting equation (9), and using the fact that  $P_0(a_{\text{H}}) = 1 - \sum_{n=1,\dots,N} P_n(a_{\text{H}})$ , the chemical shift can be expressed in terms of

the site-specific degrees of protonation as the linear combination

$$\delta_{\alpha}(a_{\text{H}}) = \delta_{\alpha 0} + \sum_{i=1, \dots, N} \delta_{\alpha i} \theta_i(a_{\text{H}}) \quad (11)$$

where  $\delta_{\alpha 0} = \tilde{\delta}_{\alpha 0}$  and  $\delta_{\alpha i} = \sum_{n=1, \dots, N} (\mathbf{A}^{-1})_{ni} (\tilde{\delta}_{\alpha n} - \tilde{\delta}_{\alpha 0})$  can be regarded as the chemical shifts of each protonated site. A significant advantage in using eqn. (11) instead of (8) is that the symmetry properties of the molecule can be incorporated in the fitting process by assuming equal  $\theta_i(a_{\text{H}})$  values for equivalent ionizable sites, as it is the case of the first and third carboxylic groups of citric acid. If chemical shift data *versus* pH are available for one or more reporter nucleus  $\alpha$ ,  $\delta_{\alpha 0}$  and the cluster parameters can be experimentally obtained using eqns. (9-11) together with the symmetry properties of the molecule.

## ***Ab initio* calculations**

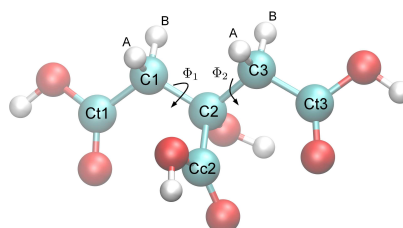
In this work, both roto-microstate and microstate probabilities are evaluated by *ab initio* calculations. The fully protonated form of the citric acid was first prepared in the extended conformation (Fig. 1a). The methylene carbons of the chain are denoted as  $C_1$ ,  $C_2$  and  $C_3$ , each of them bound to a carboxylic group, denoted as  $C_{t1}, C_{t3}$  (terminal carboxylic groups) and  $C_{c2}$  (central carboxylic group). In addition,  $C_2$  is also bound to a hydroxyl group. These three carbons form the skeleton of the molecule. The rotational state of  $C_1 - C_2$  and  $C_2 - C_3$  bonds, referred to as “central bonds”, are characterized by two dihedral angles,  $\phi_1$  and  $\phi_2$ . The rotation of each central bond results in three possible minima: *trans* (or, for abbreviation *t*, with  $\phi_i \simeq 0$ ); *gauche*<sup>+</sup> (*g*<sup>+</sup> with  $\phi_i \simeq 120$ ); and *gauche*<sup>−</sup> (*g*<sup>−</sup> with  $\phi_i \simeq -120$ ). The central bonds are considered to be in the *trans* state when the molecule is fully extended. We have chosen Flory criteria to assign the sign of the rotations, according to which the state *g*<sup>+</sup> of  $\phi_i$  is obtained by rotating the pendant groups of the carbon  $C_i$  in the counterclockwise direction<sup>41</sup> (or, in an equivalent way, rotating the pendant groups of the carbon  $C_{i+1}$  in the clockwise direction). The possible conformational states of the central bonds are represented in Fig. 1b. For each rotational state of the central bonds, there are

various possible dispositions of the carboxylic groups, which have also been explored. Because of their important contribution to the conformational stability, special attention has been paid for the viability of establishing intramolecular hydroxyl-carboxyl and carboxyl-carboxyl hydrogen bonds. With these criteria at hand, conformations were generated in a systematic way. The geometry of the conformers was optimized at MP2 level with 6-311++G(d,p) basis set using the SMD water model. All the *ab initio* calculations were carried out with Gaussian 09.<sup>42</sup> Previous studies<sup>43–45</sup> have shown the accuracy and performance of the SMD implicit model for calculations of solvation free energies. Frequency calculations in solution were performed for all the structures to confirm the absence of imaginary vibrational modes. The calculation of the vibrational contribution to the free energy in solution is required for cases where liquid and gas-phase structures differ appreciably or when stationary points present in solution do not exist in gas phase.<sup>46</sup> Free energies and roto-microstate probabilities were obtained by performing statistical thermodynamics calculations at 25°C, the temperature at which Moore *et al.* NMR experiments were performed. Microstate probabilities were calculated using eqns. (3) and (4). The same approach has been used for the rest of macrostates: the mono-ionized, di-ionized and fully ionized forms.

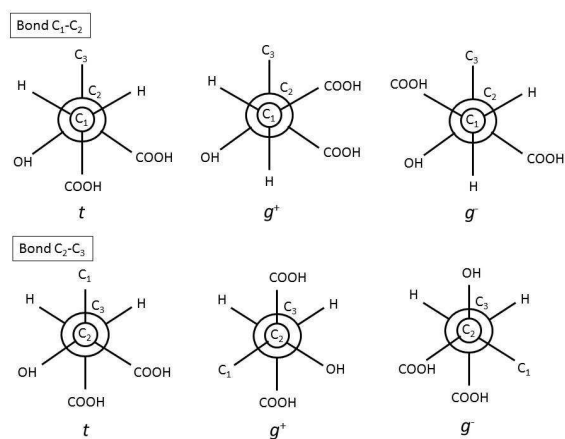
## NMR titration of citric acid

Citric acid has two non-equivalent methylene protons ( $\alpha = A, B$ ) corresponding to the two peaks of the NMR spectrum (Fig. 3). Both the separation of the peaks  $\Delta = \delta_A - \delta_B$  and the average position  $\delta = (\delta_A + \delta_B)/2$  are pH-dependent. Moreover, the two peaks are split by an amount  $J$ , due to spin-spin coupling between the two hydrogen atoms. The pH-dependence of  $\delta$ ,  $\Delta$  and  $J$  is shown in Fig. 4. High resolution  $\delta$ ,  $\Delta$  and  $J$  versus pH data were obtained by Moore and Sillerud work.<sup>24</sup> The measurements were performed at 400 MHz and TMS was used as chemical shift reference. For citric acid ( $N = 3$ ),  $\delta$  can be expressed by means of eqn. (11) as

$$\delta = \delta_0 + \delta_{13}\theta_1(a_H) + \delta_2\theta_2(a_H) \quad (12)$$



a)



b)

Figure 1: a) Representation of the fully extended citric acid molecule, showing the two main dihedral angles responsible for the conformational variability. Carbon atoms are depicted in blue, oxygen atoms in red and hydrogen atoms in white. The numbering of the carboxylic groups and the hydrogens (A and B) resolved in NMR experiments are also indicated. b) Representation of the three conformational minima obtained from the rotation of the  $C_1 - C_2$  and  $C_2 - C_3$  central bonds. The central bonds are considered to be in the *trans* ( $t$ ) state when the molecule is fully extended. We have chosen Flory criteria to assign the sign of the rotations, according to this the state *gauge+* ( $g^+$ ) of  $\phi_i$  is obtained by rotating the pendant groups of the carbon  $C_i$  in the counterclockwise direction

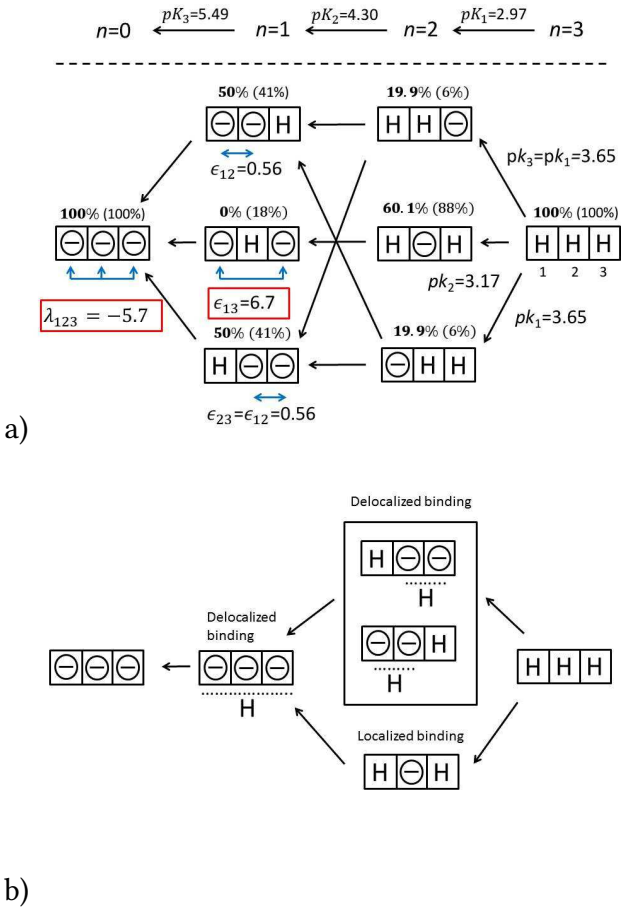


Figure 2: a) Reaction scheme of citric acid assuming fully localized proton binding.  $K_i$  are the macroscopic dissociation constants, calculated using the cluster parameters via eqns. 19. The fitted cluster parameters are reported:  $k_1$  and  $k_2$  are the dissociation microscopic constants of the terminal and central carboxylic groups, respectively.  $\epsilon_{12}$  and  $\epsilon_{13}$  represent the interaction energies between terminal-central and terminal-terminal dissociated carboxylic groups. Finally,  $\lambda_{123}$  accounts for the triplet interaction between the three dissociated carboxylic groups. The resulting microscopic probabilities are also shown (bold letters), and compared with *ab initio* calculations (between brackets). b) Reaction scheme considering the possibility of delocalized bindings. The roto-microspecies resulting from *ab initio* calculations are now classified into two groups: the ones which present delocalized proton binding and the ones which do not. Note that in the di-ionized form the proton can be exchanged among the three carboxylic groups, so that one micro-species is detected in the NMR time scale. In the mono-ionized form, however, both micro-states presenting localized and delocalized proton binding are possible.

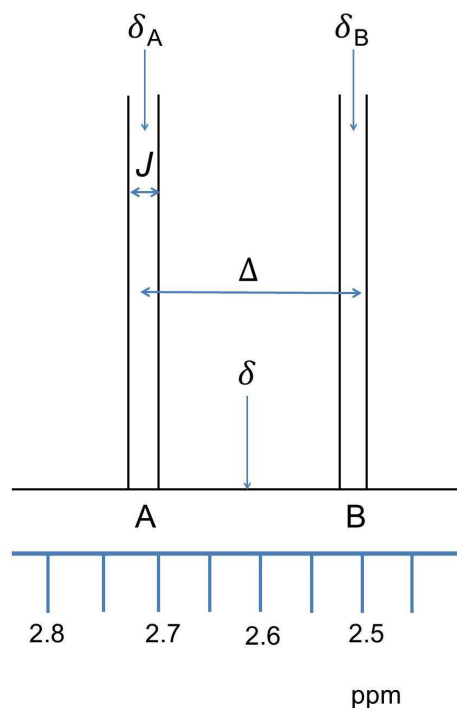


Figure 3: Sketch of the NMR citric acid spectrum corresponding to the two sets of non-equivalent methylene hydrogen atoms, A and B. The two peaks show spin-spin AB coupling, characterized by a coupling constant  $J$ . The position of the two peaks and the spin-spin coupling constant  $J$  are pH dependent. As a consequence, the average position of the two peaks ( $\delta = (\delta_A + \delta_B) / 2$ ) and the separation between the two peaks ( $\Delta = \delta_A - \delta_B$ ), reported by Moore and Sillerud<sup>24</sup> are also pH dependent. The assignment of each peak to a particular hydrogen atom, A or B, is only illustrative.

where  $\delta_i \equiv (\delta_{Ai} + \delta_{Bi})/2$ , ( $i = 0, 1, 2, 3$ ). The symmetry property  $\theta_1 = \theta_3$  has been used in (11) since the terminal carboxylic groups are equivalent.  $\delta_{13} = \delta_1 + \delta_3$ , and it accounts for the influence of both terminal carboxylic groups in the chemical shift.  $\delta_0$ , the average chemical shift for the fully deprotonated molecule, does not need to be fitted, since it can be obtained from the plateau of the experimental  $\delta$  versus pH data in the region of high pH values, where the molecule is fully deprotonated (Fig. 4a). Moreover,  $\delta_0$ ,  $\delta_{13}$  and  $\delta_2$  are not independent, since at low enough pH values, the molecule is fully protonated ( $\theta_i(a_H \rightarrow \infty) = 1$ ) giving the relationship  $\delta_0 + \delta_{13} + \delta_2 = \delta_P$ , where  $\delta_P$  is the average chemical shift for the fully protonated molecule. The value of  $\delta_P$  is obtained from the plateau of the  $\delta$  versus pH data at low pH values. Then  $\delta$  (from eqn.12) reads

$$\delta = \delta_0 + \delta_{13}\theta_1 + (\delta_P - \delta_0 - \delta_{13})\theta_2 \quad (13)$$

where  $\delta_{13}$  is the only shift parameter to be fitted. In the same way,  $\Delta$  can be expressed as

$$\Delta = \Delta_0 + \Delta_{13}\theta_1 + (\Delta_P - \Delta_0 - \Delta_{13})\theta_2 \quad (14)$$

and

$$J = J_0 + J_{13}\theta_1 + (J_P - J_0 - J_{13})\theta_2 \quad (15)$$

The validity of Eqns. (13-15) relies in the fact that  $\delta$ ,  $\Delta$  and  $J$  are linear combinations of the chemical shifts of the four individual peaks, all of them fulfilling (11). The pH-dependence of  $\delta$ ,  $\Delta$  and  $J$  comes from  $\theta_1$  and  $\theta_2$ .  $\delta_0$ ,  $\Delta_0$  and  $J_0$  are obtained from the plateau of the Figs. 4a, 4b and 4c at high enough pH values (2.42 ppm, 56.8 Hz and 15.22 Hz, respectively), while  $\delta_P$ ,  $\Delta_P$  and  $J_P$  are obtained from the plateau at low pH values (2.81 ppm, 74.5 Hz and 15.86 Hz, respectively).  $\delta_{13}$ ,  $\Delta_{13}$  and  $J_{13}$  are adjustable parameters. The chemical shifts corresponding to the individual peaks A and B,  $\delta_A$  and  $\delta_B$  are related with  $\delta$  and  $\Delta$  by

$$\delta_A = \delta - \frac{\Delta}{2f} ; \delta_B = \delta + \frac{\Delta}{2f} \quad (16)$$



where  $f$  is the spectrometer frequency. We preferred, however, to fit the original data  $\delta$ ,  $\Delta$  and  $J$  rather than  $\delta_A$ ,  $\delta_B$  and  $J$  since the peculiar shape of the curve  $\Delta$  *versus* pH can show in a clearer way the quality of the fitting.

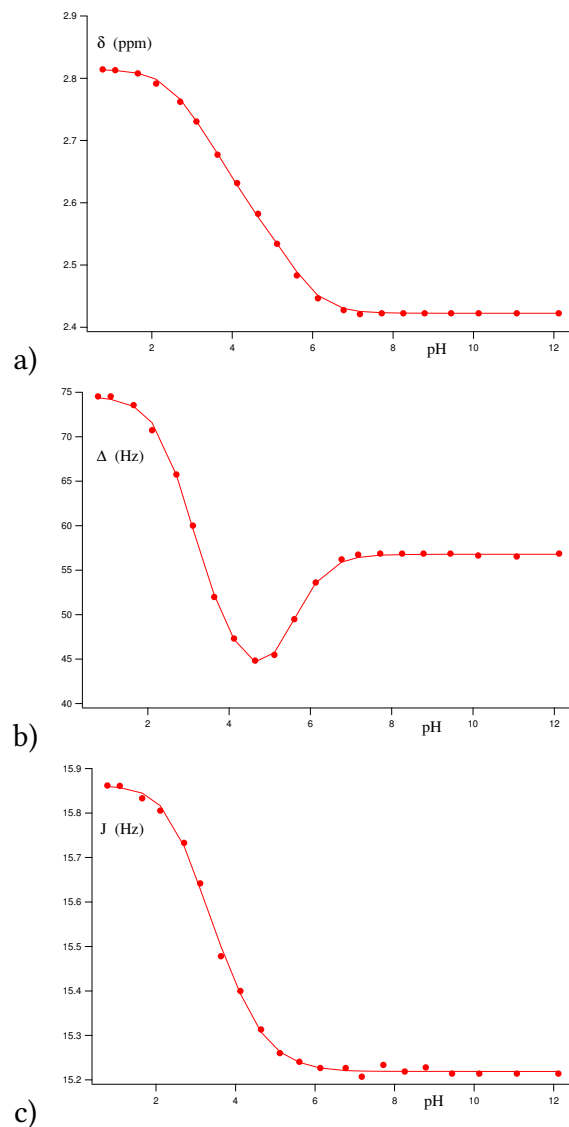


Figure 4: Markers: experimental average chemical shift ( $\delta$ ), separation between the center of the peaks ( $\Delta$ ) and spin-spin coupling ( $J$ ) *versus* pH, taken from ref.<sup>24</sup> Continuous lines: best fitted curve obtained from eqns. (13-15) together with eqns. (18-20)

On the other hand, using (9-10),  $\theta_1$  and  $\theta_2$  can be expressed in terms of the macroscopic probabilities as

$$\begin{cases} \theta_1(a_H) = \pi_1(H\Theta\Theta)P_1(a_H) + (\pi_2(H\Theta H) + \pi_2(HH\Theta))P_2(a_H) + P_3(a_H) \\ \theta_2(a_H) = \pi_1(\Theta H\Theta)P_1(a_H) + (\pi_2(HH\Theta) + \pi_2(\Theta HH))P_2(a_H) + P_3(a_H) \end{cases} \quad (17)$$

where H and  $\Theta$  represent a protonated ( $-\text{COOH}$ ) and a deprotonated ( $-\text{COO}^-$ ) site, respectively. The conditional probabilities are not independent since they are constrained by symmetry:  $\pi_1(\Theta\Theta H) = \pi_1(H\Theta\Theta)$ ,  $\pi_2(\Theta HH) = \pi_2(HH\Theta)$  and by the normalization conditions  $\pi_1(\Theta H\Theta) + \pi_1(H\Theta\Theta) + \pi_1(\Theta\Theta H) = 1$  and  $\pi_2(H\Theta H) + \pi_2(HH\Theta) + \pi_2(\Theta HH) = 1$ . Therefore, there are only 2 independent conditional probabilities, for instance,  $\pi_1(H\Theta\Theta)$  and  $\pi_2(HH\Theta)$ , and (17) can be rewritten as

$$\begin{cases} \theta_1(a_H) = \pi_1(H\Theta\Theta)P_1(a_H) + (1 - \pi_2(HH\Theta))P_2(a_H) + P_3(a_H) \\ \theta_2(a_H) = (1 - 2\pi_1(H\Theta\Theta))P_1(a_H) + 2\pi_2(HH\Theta)P_2(a_H) + P_3(a_H) \end{cases} \quad (18)$$

Eqns. (17) can be fully expressed in terms of the cluster parameters. The macrostate probabilities  $P_i(a_H)$  depend on the macroconstants through eqns. 2, which, after some elementary calculations, can be related to the microscopic binding parameters of the cluster expansion (6) as follows

$$\begin{cases} K_1 = 2k_1 + k_2 \\ K_2 = \frac{k_1(2k_2u + k_1v)}{2k_1 + k_2} \\ K_3 = \frac{k_1k_2u^2vw}{2k_2u + k_1v} \end{cases} \quad (19)$$

with  $\log u = -\epsilon_{12}$ ,  $\log v = -\epsilon_{13}$  and  $\log w = -\lambda_{123}$ . The microstate probabilities are given by:

$$\pi_1(H\Theta\Theta) = \frac{k_2u}{2k_2u + k_1v} ; \quad \pi_2(HH\Theta) = \frac{k_1}{2k_1 + k_2} \quad (20)$$

In summary, equations (13-15) and (18-20) contain eight parameters to be simultaneously fitted to the experimental chemical shifts: three chemical shift parameters ( $\delta_{13}$ ,  $\Delta_{13}$  and  $J_{13}$ ), two

microscopic dissociation constants ( $k_1$  and  $k_2$ ) and three interaction energies ( $\epsilon_{12}$ ,  $\epsilon_{13}$  and  $\lambda_{123}$ ). The fitting procedure has been performed using standard nonlinear regression routines of MATLAB.<sup>47</sup> The robustness of the fitting has been checked by means of Monte Carlo simulation of synthetic data, which provides accurate estimates of the parameter confidence intervals without requiring any assumption about the probability distribution of the parameters, or about the linearity of the fitted mathematical model.<sup>48,49</sup> In this method, an initial best fitting is performed using the original data set. The average deviation between the best-fitted curve and the experimental data is taken as an estimation of the variance  $\sigma$  of the error probability distribution, which is assumed to be Gaussian. Then a new set of synthetic data set is generated by adding random errors to the experimental data. A fitting process is performed to the new data set and a new set of best-fitted parameters is obtained. The process is repeated a sufficiently large number of times (thousands in this work), and the resulting values for each particular fitted parameters are listed in numerical order. The confidence intervals of 95%, the ones reported in this work, are obtained by removing from the ordered list those values corresponding to the 0.5% of the upper and lower tail. Since the probability distribution of the parameters does not need to be Gaussian, the confidence intervals do not need be strictly symmetric. Hence both the upper and lower bounds of the limiting intervals are reported (see Tables 1 and 2). Confidence intervals for functions of the fitted parameters, such as microstate probabilities or macroscopic dissociation constants can be obtained by the same procedure.

The found NMR parameters in eqns. (12-15) are shown in Table 1. The best-fitted parameters  $\delta_{13}$ ,  $\Delta_{13}$  and  $J_{13}$  are reported within 95% confidence intervals. If the fitting is performed only to  $\delta$  and  $\Delta$  data (Figs. 4a and 4b), without using the spin-spin coupling data (Fig. 4c) the results do not change significantly. This fact supports the consistency of the obtained values.  $\delta$ ,  $\Delta$  and  $J$  for the fully protonated ( $\delta_P$ ,  $\Delta_P$  and  $J_P$ ) and the fully deprotonated ( $\delta_0$ ,  $\Delta_0$  and  $J_0$ ) molecule were not fitted but obtained from the plateaus of figs. 4 at low and high pH-values, respectively.

The cluster parameters are shown in Table 2, together with the ones calculated from the results reported in refs.<sup>14,15</sup> These works make use of selective methyl substitution of the carboxylic

Table 1: Values of the  $\delta_i$ ,  $\Delta_i$  and  $J_i$  NMR shift parameters for citric acid defined in eqns. 13-15. The NMR shifts corresponding to the fully deprotonated ( $i=0$ ) or fully protonated ( $i=P$ ) have been obtained from the plateau of the curves of Fig. 4 at high and low pH-values, respectively.  $\delta_{13}$ ,  $\Delta_{13}$  and  $J_{13}$  have been fitted to the three experimental curves in Fig. 4 simultaneously. The 95% confidence interval has been estimated using a Monte Carlo method consisting in the generation of synthetic data with random Gaussian errors with variance  $2\sigma$ .  $\delta_2$ ,  $\Delta_2$  and  $J_2$  have been calculated using  $\delta_2 = \delta_P - \delta_0 - \delta_{13}$ ,  $\Delta_2 = \Delta_P - \Delta_0 - \Delta_{13}$  and  $J_2 = J_P - J_0 - J_{13}$ , respectively.

$i$	13	0	P	2
$\delta_i$ (ppm)	0.28 (+0.02/-0.02)	2.42	2.81	0.11
$J_i$ (Hz)	0.07 (+0.05/-0.06)	15.21	15.81	0.53
$\Delta_i$ (Hz)	-31.7 (+1.1/-1.2)	56.7	74.5	49.4

groups to deduce the complete ionization scheme of citric acid. The accuracy of these methods relies on the assumption that the non-blocked groups have the same ionization properties as the original molecule. Under this assumption, potentiometric titration of the modified molecules allows determining two of the cluster parameters (for instance  $pk_1$  and  $\epsilon_{12}$ ). Provided that the macroscopic dissociation constants are known, the restant three cluster parameters can be obtained by solving the three eqns. 19. In the case of Martin work,<sup>14</sup> we have used the macroscopic constants at  $I=0.1$  M reported in Robertis *et al.*,<sup>50</sup> instead of those at infinite dilution,<sup>6</sup> which leads to important discrepancies with Pearce *et al.* results.<sup>15</sup> As previously pointed out in ref.,<sup>15</sup> Martin does not state the concentrations of ester used in the titrations, but these are supposed to be high enough to significantly contribute to the ionic strength. This consideration seems confirmed by the good agreement between both works when the macroscopic dissociation constants are not taken at infinite dilution (Table 2).

The microscopic dissociation  $pk$ -values,  $pk_1 = 3.65$ ,  $pk_2 = 3.17$ , and the nearest neighbour interaction energy,  $\epsilon_{12} = 0.56$ , are, within a narrow confidence interval, very similar to those obtained from refs.<sup>14,15</sup> at 0.1M of KCl solution. However, the next-nearest interaction parameter  $\epsilon_{13}$  and the triplet interaction parameter,  $\lambda_{123}$  clearly differ from those obtained from the previous works. The obtained values,  $\epsilon_{13} = 6.68$  and  $\lambda_{123} = -5.65$ , represent enormous repulsive (9.1 kcal/mol) and attractive energies (-7.7 kcal/mol), respectively. Monte Carlo error analysis shows that this behaviour is very robust within the whole confidence intervals, reported in Table 2

indicate. In order to understand how unrealistic these values are, we can compare them with the interaction energies of nearest-neighbour sites of other molecules whose ionizable groups are separated by three bonds. For instance, for poly(maleic) and poly(fumaric) acids, with interactions regarded as particularly strong,  $\epsilon_{12} \simeq 1.5 - 3.9$ , depending on the background electrolyte and the bond tacticity; for succinic acid  $\epsilon_{12} \simeq 0.8$ ;<sup>51</sup> for poly(ethylene)imine,  $\epsilon_{12} \simeq 2$  and  $\lambda_{123} \simeq 0.45$ .<sup>52</sup> Moreover, it is important to highlight the strange fact that the interaction energy between terminal charged groups  $\epsilon_{13}$  is much larger than that between a terminal and the central group. This is counterintuitive and very difficult to justify in physical terms, since when the two terminal groups are charged, the molecule can accommodate the conformation to get a larger distance between the charged groups than when a terminal and a central group are charged.

Once the cluster parameters are determined, the macroscopic dissociation constants can be calculated using eqns. 19. Interestingly, despite the unreliable values of  $\epsilon_{13}$  and  $\lambda_{123}$ , the macroscopic constants, calculated using eqns. 19, are in good agreement with the values reported in previous literature, as shown in Table 3. Especially remarkable is the agreement with the values reported by Robertis *et al.*<sup>50</sup> (with  $I=0.1\text{M}$  and  $I=0.5\text{ M}$ ), Crea *et al.*<sup>8</sup> (with  $I=0.5\text{M}$ ), and Bénézech *et al.*<sup>7</sup> ( $I=0.1\text{m}$  and  $I=0.5\text{ m}$ ). They also agree with other experimental studies,<sup>9,53,54</sup> and calculations using the Pitzer theory.<sup>8,50,55</sup> In contrast, they differ with the values obtained at infinite dilution.<sup>6</sup> This discrepancy is expected since the NMR experiments were done at a sodium citrate concentration  $0.1\text{ M}$ , so that the minimum value of the ionic strength in the titration experiment was  $I=0.3\text{ M}$ .

In summary, in analyzing the pH-dependence of the NMR chemical shifts by means of the SB model, *i.e.*, assuming fully localized proton binding, we arrive to contradictory results: on the one side, we obtain good values for the macroscopic constants, in agreement with those obtained previously with potentiometry. But, on the other side, we get nonphysical values for the interaction energies. This intriguing fact is also reflected in the obtained microstate probabilities, which are shown in Table 4. They clearly differ from those obtained from refs.<sup>14</sup> and.<sup>15</sup> These works predict a significant presence of the symmetric di-ionized microstate ( $\Theta\text{H}\Theta$ ) (around 10-15%).

As will be shown in section 4, *ab initio* computations assuming localized proton binding also lead to the same conclusion. Conversely, our analysis shows that, within a very narrow confidence interval, this microstate is virtually nonexistent, so that only the asymmetric di-ionized microstate is present. Algebraic arguments, complementary to the Monte Carlo test, can also be used to justify the robustness of this result. The analysis, outlined in the supporting information, shows the existence of a unique minimum in the fitting process. The nonexistence of the di-ionized symmetric microstate is equivalent to a huge repulsive energy between the terminal carboxylates. In the same way, this huge repulsive energy must be compensated by an enormous triplet attractive energy to produce the tri-ionized state. Actually, it can be shown that the confidence interval for the sum of  $\epsilon_{13}$  and  $\lambda_{123}$  is very narrow, so that triplet interactions always compensate the terminal-terminal carboxylate repulsion.

In this work we propose that this result, clearly an artifact, could be the consequence of trying to analyze the NMR shifts assuming fully localized proton. Aiming at clarifying this point, in the next section a systematic scan of the most stable conformations of citric acid is performed. As will be shown, some of this conformations allow proton exchange between carboxylic groups without energetic barrier, so that the basic assumption of the SB model, i.e., the assignment a carboxylic group to a bound proton, is no longer suitable to understand the ionization properties of citric acid. It is worth to note that proton delocalization can not take place in the experiments which use selective methylation of the carboxylic groups. This would explain the observed discrepancies with the results obtained from NMR titrations.

Table 2: Cluster parameters of citric acid at 25°C obtained by fitting the SB model (eqns. 18-19) to NMR titration data taken from ref.<sup>24</sup> The best-fitted values are given within 95% confidence intervals, which are reported between brackets. The values calculated from refs.<sup>14</sup> and,<sup>15</sup> which use selective methylation of the carboxylic groups, are also reported.

	$pk_1$	$pk_2$	$\epsilon_{12}$	$\epsilon_{13}$	$\lambda_{123}$
fitted from NMR titration	3.65	3.17	0.56	6.68	-5.65
$I \geq 0.3$ M	(+0.04/ -0.03)	(+0.03/ -0.02)	(+0.08/ -0.08)	(+4.02/ -4.63)	(+4.67/ -4.07)
Martin <sup>14</sup> $I \sim 0.1$ M	3.85	3.35	0.73	0.54	-0.28
Pearce <i>et al.</i> <sup>15</sup> $I = 0.1$ M	3.79	3.02	0.78	0.62	-0.15

Table 3: Macroscopic  $pK$ -values of citric acid at 25°C calculated using the cluster parameters obtained from NMR titrations (Table 2) and eqns. (19), together with values previously reported in the literature, determined by potentiometry.

	$pK_1$	$pK_2$	$pK_3$
from NMR titrations (Table 2)	2.95	4.13	5.54
using eqns. 19, $I \geq 0.3$ M	(+0.03/-0.03)	(+0.05/-0.05)	(+0.05/-0.05)
Bates <i>et al.</i> <sup>6</sup> $I \sim 0$ M	$3.13 \pm 0.02$	$4.76 \pm 0.01$	$6.40 \pm 0.01$
Pearce <i>et al.</i> <sup>15</sup> $I \sim 0$ M	3.04	4.70	5.77
Robertis <i>et al.</i> <sup>50</sup> $I=0.1$ M	2.92	4.35	5.77-5.78
Bénezeth <i>et al.</i> <sup>7</sup> $I=0.1$ m	$2.90 \pm 0.01$	$4.34 \pm 0.01$	$5.68 \pm 0.02$
Robertis <i>et al.</i> <sup>50</sup> $I=0.5$ M	2.82	4.13-4.15	5.41-5.44
Crea <i>et al.</i> <sup>8</sup> $I=0.5$ M	2.81-2.82	4.11-4.15	5.31-5.44
Bénezeth <i>et al.</i> <sup>7</sup> $I=0.5$ m	$2.77 \pm 0.01$	$4.12 \pm 0.02$	$5.27 \pm 0.04$

Table 4: Microstate probabilities assuming fully localized proton binding at 25°C, calculated replacing the found cluster parameters (Table 2) in eqns. (20), together with values previously reported in the literature, determined by potentiometry.

	$\pi_2(\text{H}\Theta\text{H})$	$\pi_2(\text{HH}\Theta)=$ $\pi_2(\Theta\text{HH})$	$\pi_1(\Theta\text{H}\Theta)$	$\pi_1(\text{H}\Theta\Theta)=$ $\pi_1(\Theta\Theta\text{H})$
fitted from NMR titration	60.1%	19.9%	0.0%	50.0%
$I \geq 0.3$ M	(+0.5/ -0.0)	(+0.2/ -0.3)	(+0.4/ -0.0)	(+0.0/ -0.2)
<i>ab initio</i> calculations $I = 0$ M	88%	6%	18%	41%
Martin <sup>14</sup> $I \sim 0.1$ M	77%	11.5%	10.6%	44.7%
Pearce <i>et al.</i> <sup>15</sup> $I \sim 0$ M (exp)	79%	10.5%	11%	44.5%
Pearce <i>et al.</i> <sup>15</sup> $I = 0.1$ M (exp)	75%	12.5%	14%	43%

## ***Ab initio* calculations and hydrogen bond formation. Localized versus delocalized proton binding**

The most stable roto-microstates of citric acid for the different degrees of ionization are shown in Table 5. They are obtained by means of *ab initio* MP2 computations with 6-311++G(d,p) basis set using SMD water model. Free energies values are reported as relative to the most stable roto-microstate of the same macrostate, which is given zero free energy. The roto-microstate probability within a specific macrostate is then calculated using the Boltzmann distribution at the same temperature of NMR experiments, 25°C. The possibility of state multiplicity due to symmetry is also taken into account in the calculations. Only roto-microstates with probability larger than 0.9% are shown in Table 5. The rest of them are reported in the supplementary information. Once a particular macrostate was fixed, the conformational search of citric acid was carried out preparing initial structures with *gauche+*, *gauche-* and *trans* conformations for  $\phi_1$  and  $\phi_2$  dihedral angles. In addition, different orientations of the hydroxyl and carboxylic groups were also sampled in a systematic way. Special attention has been paid to the formation of carboxyl-carboxyl and carboxyl-hydroxyl hydrogen bonds because of their stabilizing capacity of the conformation. The HB bond distance is also shown in Table 5. We consider that a HB is formed if the distance between oxygen atoms is less than 3 angstroms and the angle O-H-O is larger than 120 degrees. HB bonds between two carboxylic groups, one of them acting as an acceptor and the other one as a donor, are indicated in the Table by an overbar on the corresponding deprotonated (“ $\Theta$ ”) or protonated (“H”) site. The computations have firstly been performed assuming fully localized proton binding.

In Table 5, the roto-microstates are depicted by means of two letters indicating the conformations of  $\phi_1$  and  $\phi_2$  dihedral angles, followed by a number indicating the relative stability with respect to the most stable roto-microstate having the same conformation. There is also an additional sub-index expressing the ionization state (0, -1, -2, -3). So, for instance,  $tg^{+1}_{-2}$  represents a di-ionized roto-microstate with  $\phi_1$  in *trans* state,  $\phi_2$  in *gauge+* state, and with the most stable



disposition in the  $tg^+$  conformation.

For the fully protonated molecule,  $C_6H_8O_7$ , three minima were found with significant population: one for  $tt$  and two for  $tg^+$ .  $tt1_0$  corresponds to the fully extended molecule being the most stable one (88%). It is not completely symmetric ( $\phi_1 \neq -\phi_2$ ), since the hydroxyl group is oriented towards one of the terminal carboxylic groups while the other terminal carboxylic group is oriented differently. Roto-microstates  $tg^+1_0$  and  $tg^+2_0$ , with very similar free energies, are much less abundant (7.0% and 4.5%, respectively). In  $tg^+1_0$  and  $tg^+2_0$  the hydroxyl group points towards the *gauche* and the *trans* part of the molecule, respectively.

In the case of the fully deprotonated citrate anion,  $C_6H_5O_7^{-3}$ , the most stable conformation is  $tt1_{-3}$  followed by  $tg^+1_{-3}$  (0.4 kcal/mol). Thus,  $tt1_{-3}$  has the greatest abundant of the fully deprotonated form, with 62 %, followed by 33% of abundance for  $tg^+1_{-3}$

For the hydrogen citrate anion,  $C_6H_7O_7^{-1}$ , the more stable conformers are of the type  $tt$  (with five dispositions),  $tg^+$  (one disposition) and  $g^-g^-$  (five dispositions). The most stables structures, depicted in Fig. 5a and 5b, correspond to  $tt1_{-1}$  (with probability of 23%) and  $g^-g^-1_{-1}$  (48%) conformations, with the same free energy. Assuming fully localized binding, in both cases the central carboxylic group is ionized. In the case of  $g^-g^-1_{-1}$ , a HB bond is formed between the central deionized carboxyl and the terminal one, which clearly stabilizes this conformation. Together with the roto-microstates  $tt2_{-1}$ ,  $tt3_{-1}$ ,  $tt4_{-1}$ ,  $tt5_{-1}$  and  $tg^+1_{-1}$ , all of them with an ionized central carboxyl, the resulting probability for the symmetric charged form is 88%. This is in agreement with refs.,<sup>14,15</sup> where the micro-state probabilities were obtained using selective methylation of the carboxylic groups, a technique for which HB are destroyed and therefore delocalized proton binding is unavoidable. The comparison with the micro-state probabilities obtained from these experiments is shown in Table 4. The agreement is remarkable taking into account that water has been included implicitly. Models with implicit solvent have revealed to be very useful in, for instance, differential solvation in chemical reactions.<sup>56</sup> More complete models including explicit water could lead to a more complete picture and are object of current research.

For the di-ionized species,  $C_6H_6O_7^{-2}$ , again reasonable agreement between our *ab initio* calcu-

lations and refs.<sup>14,15</sup> is obtained. However, the micro-state probabilities are in total contradiction with the NMR titration experiments. In our opinion, this discrepancy is, as revealed by Monte-Carlo statistical analysis, robust enough to inspect alternatives to the standard SB model.

The picture changes considerably if we accept the possibility that at least some of the formed HB allow the proton to be delocalized between two charged carboxylic groups. This means that, in reality, one should consider three possible states for the proton: localized in the central carboxyl, localized in the terminal carboxyl, and delocalized proton between two carboxylic groups. The existence of this 'third' species was proposed by Kawaguchi *et al.*,<sup>31</sup> who suggested the existence of shared protons between consecutive carboxylic group to explain the pH-dependence of UV and IR spectra of poly(fumaric) and poly(maleic) acids. Let us firstly classify the roto-microstates of  $C_6H_7O_7^{-1}$  into two groups: those which contain a HB bond between carboxylic groups and those which do not.  $tt1_{-1}$  (with probability 23%),  $tt5_{-1}$  (1.1%) and  $tg^+1_{-1}$  (6.1%), which all represent 30.2% of the roto-microstates, do not present any kind of HB. On the contrary,  $g^-g^-1_{-1}$ ,  $g^-g^-2_{-1}$ ,  $tt2_{-1}$ ,  $tt3_{-1}$  and  $tt4_{-1}$  exhibit HB between the central and the terminal carboxylic groups. In the case of  $tt2_{-1}$ , moreover, the three carboxylic groups participate in hydrogen bonds, with the two terminal carboxylic groups acting as hydrogen donors and the central one as an acceptor. They are thus candidates to present proton delocalization.

Let's explore the possibility of proton delocalization by means of *ab initio* calculations. Firstly note that in the conformations  $g^-g^-1_{-1}$  and  $g^-g^-2_{-1}$ , all the atoms have practically the same coordinates, except that of the acid hydrogen, as can be observed in Figs. 5b and 5c. A shift in the acid hydrogen position of about 0.3 angstroms makes to protonate a lateral ( $g^-g^-1_{-1}$ ) or the central ( $g^-g^-2_{-1}$ ) carboxylic group. In order to analyze the free energy profile of the transition, additional calculations were performed by locating the acid hydrogen atom at intermediate positions between both carboxylic groups. The obtained results are shown in Fig. 6a, where we have plotted energy *versus* distance to the center between the two oxygens. The resulting curve indicates that the transition between  $g^-g^-1_{-1}$  and  $g^-g^-2_{-1}$  is produced with a small energetic barrier of about 0.2 kcal/mol. The most stable conformer is obtained when the proton is

in between these two conformational minima. We can observe, however, that the localization of the proton in this transformation is not symmetric, being more similar to the  $g^-g^-1_{-1}$  minima. There are also species without HB between carboxylic groups:  $tt1_{-1}$ ,  $tt5_{-1}$  and  $tg^+1_{-1}$ , all of them corresponding to the symmetric form, for which the central carboxyl is ionized. Finally, in the roto-microstates  $tt2_{-1}$ ,  $tt3_{-1}$  and  $tt4_{-1}$  the central carboxylic group is ionized and the lateral carboxyl groups are protonated, but in this case forming a HB without proton delocalization. It is worth to note, that those HB that are localized are those with the greatest HB distances ( $> 1.48$  Å), while those that are delocalized (SSHB) have the shortest ones ( $< 1.48$  Å) (Table 5). In summary, the roto-microstates corresponding to the dihydrogen-citrate ion can be classified into two groups: the symmetric form with localized terminal protons and the one with delocalized protons. These values are very similar to the results obtained from NMR titrations, if the micro-species are re-interpreted as symmetric with terminal localized protons *versus* asymmetric with delocalized proton. The new interpretation of the reaction scheme is outlined in Fig. 2b. Unfortunately, when proton delocalization taken into account, some information, such as the entropy associated to proton delocalization, is lacking in order to rigorously estimate the probabilities of localized and delocalized states.

In the case of the hydrogen-citrate anion  $C_6H_6O_7^{-2}$ , proton delocalization could play an even more important role. All the most stable conformations reported in Table 5 can be transformed one into other either by a conformational change, or by transference of delocalized a proton through a HB without energetic barrier (SSLB). This means that at the NMR time scale, all the roto-microstates behave on average as a unique microstate, which is consistent with the surprising 100% probability for only one microstate. A large part of the population (61%) is distributed between only two minima: the  $g^-g^-1_{-2}$  and  $g^-g^-2_{-2}$  asymmetrical charged conformers (Figs. 5d and 5e). The only difference between these two conformers is the transference of a proton between the two terminal carboxyl groups. Calculations fixing the acid hydrogen at intermediate positions indicates that this transition is performed without energetic barrier, as shown in Fig. 6b. Actually, the curve energy *versus* distance to the center of the oxygens is in this case very

symmetric and exhibits a minimum of about -2 kcal/mol. A similar situation is found for the pairs of conformers  $g^-g^-3_{-2}$  vs  $g^-g^-4_{-2}$  and  $tg^+1_{-2}$  vs  $tg^+2_{-2}$ , as shown in Figs. 6c and 6d. Moreover,  $g^-g^-5_{-2}$  can be transformed into  $g^-g^-4_{-2}$  and  $tg^+2_{-2}$  into  $tg^+3_{-2}$  just by a change in the disposition of the resting carboxylic group. As a consequence, within this new picture, there would not be localized protons in the di-ionized form of citric acid, so that NMR can only detect a unique species. We have depicted this result in Fig. 2b. The large discrepancy between the speciation resulting from NMR titrations and previous works, based on methyl substitution emerges now as a natural consequence of proton delocalization. Selected blocking of the carboxylic groups would break the HB of the original molecule, so that the modified compound behaves as the original one presenting only fully localized protons. This would also explain the rather good agreement, within the limitations of the calculation technique, between “deductive” methods in refs.<sup>14</sup> and<sup>15</sup> and *ab initio* computations.

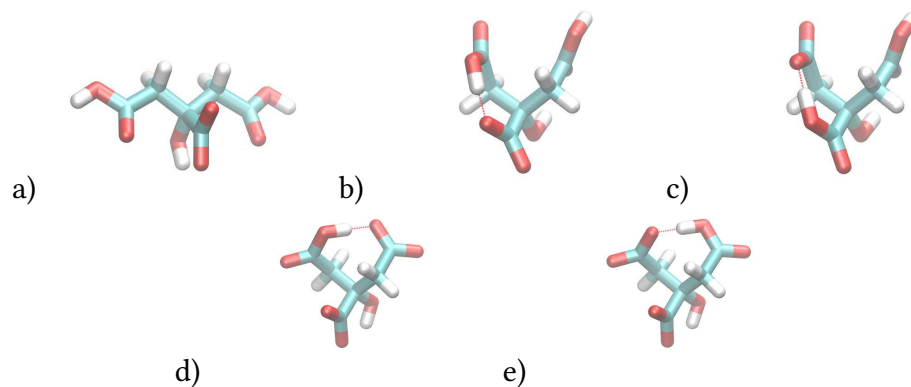


Figure 5: More stable roto-microstates of di-hydrogen citrate ( $q=-1$ ):  $tt1_{-1}$  ( $H\Theta H$ ) (a),  $g^-g^-1_{-1}$  ( $H\Theta H$ ) (b) and  $g^-g^-2_{-1}$  ( $\Theta H \Theta$ ) (c); and hydrogen citrate ( $q=-2$ ):  $g^-g^-1_{-2}$  ( $H\Theta \Theta$ ) (d) and  $g^-g^-2_{-2}$  ( $\Theta \Theta H$ ) (e). Oxygen atoms are depicted in red, carbons in green and hydrogen atoms in white.

## Conclusions

The micro-speciation of citric acid is studied by analyzing NMR titration data. When fully localized proton binding is assumed, the so-called site-binding (SB) model, microscopic dissociation constants, together with pair and triplet interaction energies between charged carboxylate groups

Table 5: Relative free energies of conformations of citrate in water obtained by MP2 calculations for all the possible degrees of ionization (0,-1,-2,-3). “H” indicates a protonated carboxylic group while “ $\Theta$ ” depicts a deprotonated carboxylic group. The participation of a carboxylic group in a intramolecular hydrogen bond between two carboxylic groups is indicated by an overbar over the corresponding “H” and “ $\Theta$ ” sites. Only conformations with a proportion greater than 0.9% are shown. Relative Free energies in water (in kcal/mol) and hydrogen bond dinstances (in angstroms). Conformational proportions were calculated using proper Boltzmann factors at 25°C. The hydrogen bond are classified into two groups: those which present localized proton binding and those who canform a SSLB-HB.

charge	conformation	protonation	$\Delta G_{rel}$ (kcal/mol)	Relative population	d(H...O) (Å)	HB type
0	<i>tt1</i> <sub>0</sub>	HHH	0.0	88%	—	—
	<i>tg</i> <sup>+</sup> <i>1</i> <sub>0</sub>	HHH	1.5	6.9%	—	—
	<i>tg</i> <sup>+</sup> <i>2</i> <sub>0</sub>	HHH	1.8	4.4%	—	—
-1	<i>tt1</i> <sub>-1</sub>	H $\Theta$ H	0.0	23%	—	—
	<i>tt2</i> <sub>-1</sub>	$\overline{\text{H}\Theta\text{H}}$	1.5	3.7%	1.50	local.
	<i>tt3</i> <sub>-1</sub>	$\overline{\text{H}\Theta\text{H}}$	1.6	3.2%	1.63	local.
	<i>tt4</i> <sub>-1</sub>	$\overline{\text{H}\Theta\text{H}}$	2.1	1.5%	2.13	local.
	<i>tt5</i> <sub>-1</sub>	H $\Theta$ H	2.2	1.1%	—	—
	<i>tg</i> <sup>+</sup> <i>1</i> <sub>-1</sub>	H $\Theta$ H	1.2	6.1%	—	—
	<i>g</i> <sup>-</sup> <i>g</i> <sup>-</sup> <i>1</i> <sub>-1</sub>	$\overline{\text{H}\Theta\text{H}}$	0.0	48%	1.46	SSLB
	<i>g</i> <sup>-</sup> <i>g</i> <sup>-</sup> <i>2</i> <sub>-1</sub>	$\overline{\Theta\text{H}\text{H}}$	0.9	10%	1.36	SSLB
-2	<i>tg</i> <sup>+</sup> <i>1</i> <sub>-2</sub>	$\Theta\text{H}\Theta$	0.8	10%	1.37	SSLB
	<i>tg</i> <sup>+</sup> <i>2</i> <sub>-2</sub>	$\Theta\overline{\Theta\text{H}}$	1.2	5.3%	1.43	SSLB
	<i>tg</i> <sup>+</sup> <i>3</i> <sub>-2</sub>	$\Theta\overline{\Theta\text{H}}$	1.5	3.1%	1.47	SSLB
	<i>g</i> <sup>-</sup> <i>g</i> <sup>-</sup> <i>1</i> <sub>-2</sub>	$\overline{\text{H}\Theta\Theta}$	0.0	37%	1.47	SSLB
	<i>g</i> <sup>-</sup> <i>g</i> <sup>-</sup> <i>2</i> <sub>-2</sub>	$\overline{\Theta\Theta\text{H}}$	0.3	24%	1.46	SSLB
	<i>g</i> <sup>-</sup> <i>g</i> <sup>-</sup> <i>3</i> <sub>-2</sub>	$\overline{\Theta\text{H}\Theta}$	0.9	8.2%	1.36	SSLB
	<i>g</i> <sup>-</sup> <i>g</i> <sup>-</sup> <i>4</i> <sub>-2</sub>	$\overline{\text{H}\Theta\Theta}$	1.0	6.7%	1.44	SSLB
	<i>g</i> <sup>-</sup> <i>g</i> <sup>-</sup> <i>5</i> <sub>-2</sub>	$\overline{\text{H}\Theta\Theta}$	1.3	4.0%	1.47	SSLB
-3	<i>tt1</i> <sub>-3</sub>	$\Theta\Theta\Theta$	0.0	62%	—	—
	<i>tg</i> <sup>+</sup> <i>1</i> <sub>-3</sub>	$\Theta\Theta\Theta$	0.4	33%	—	—
	<i>tg</i> <sup>+</sup> <i>2</i> <sub>-3</sub>	$\Theta\Theta\Theta$	1.6	3.8%	—	—

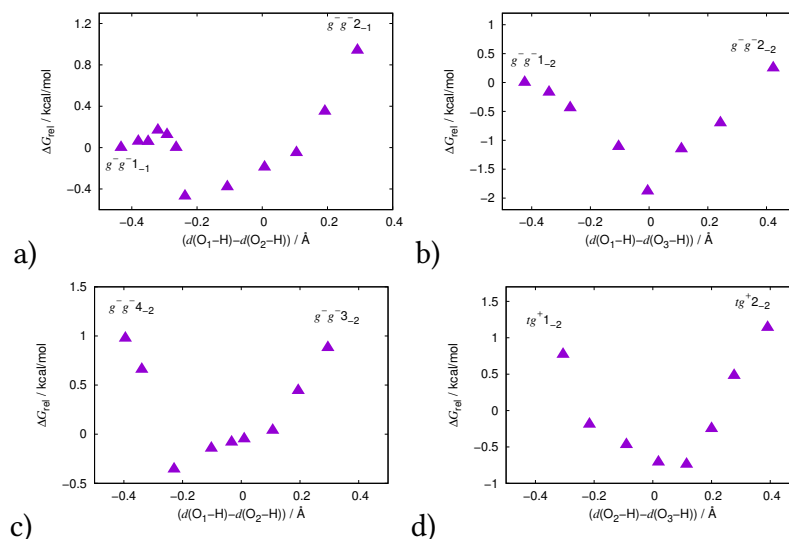


Figure 6: Free energy profiles for the intramolecular proton transfer between a)  $g^-g^-1_{-1}$  and  $g^-g^-1_{-1}$ ; b)  $g^-g^-1_{-2}$  and  $g^-g^-2_{-2}$ ; c)  $g^-g^-3_{-2}$  and  $g^-g^-4_{-2}$ ; d)  $tg^+1_{-2}$  and  $tg^+2_{-2}$  as a function of the deviation of the proton from its centered position in the hydrogen bond.

are obtained. The resulting macroscopic constants are in very good agreement with the values reported in previous literature using potentiometry. However, the found pair interaction energy between two terminal carboxylates and the triplet interaction energy are physically meaningless. The obtained interactions between two terminal charged groups are much larger than that between a terminal and the central groups. To be sure that this strange result is not an artifact produced by uncertainties in the fitting procedure, a robustness test based on random generation of synthetic data has been performed. Algebraic arguments also lead to the same conclusion.

This apparent contradiction suggested a more detailed analysis of the standard reaction scheme based on the SB model. As a consequence, full elucidation of the conformational structures for each ionization state of citric acid became necessary. Special attention has been paid for the formation of hydrogen bonds, due to their stabilizing effect. With this aim, *ab initio* MP2 calculations using the SMD polarizable continuum model for solvent were performed. When protons are constrained to be bound to a particular carboxylic group, the resulting micro-state probabilities are in reasonable agreement with the micro-speciation reported in previous works using selective blocking of the carboxylic groups. They were, however, in disagreement with the microstate

probabilities derived from the NMR titrations, specially in the case of the hydrogen-citrate ion.

A more careful analysis of the calculations suggested the possibility of proton delocalization between two or even three carboxylate groups. The computations indicate that both dihydrogen and hydrogen citrate adopt stable conformations which could present proton delocalization through low barrier HB between carboxyl groups. In the case of the hydrogen citrate ion, we find that all the stable conformations (with probability near 100%) present this possibility. This suggest a new reaction scheme including micro-species presenting delocalized proton binding. In this new scheme, NMR would only detect a unique micro-species for the hydrogen-citrate ion. This is what is concluded from the analysis of NMR titration data and would explain why when the protons are constrained to be localized, the fundamental assumption of the SB model, non-physical interaction energies are obtained. Micro-speciation of di-hydrogen citrate ion seems to be better explained if delocalized proton binding is considered for some of the roto-microstates.

Delocalized proton binding in solution is an open and controversial issue, which has been the subject of recent research. This work suggests that NMR in solution, combined with suitable micro-speciation analysis and *ab initio* calculations, could shed some more light on this important topic.

## Acknowledgement

We acknowledge the financial support from Generalitat de Catalunya (Grants 2014SGR1017, 2014SGR1132 and XrQTC). J.L.G. acknowledge the Spanish Ministry of Science and Innovation (project CTM2016-78798-C2-1-P). S.M., M.N. and F.M. acknowledge the funding of the 8SEWP - HORIZON-2020 grant "Materials Networking" (692146).

## Supporting Information Available

- An additional argument to the robustness of the found microscopic binding parameters.

- Additional figures showing free energy minimum conformations of citric acid obtained from *ab initio* optimizations in solvent for fully protonated, mono-deprotonated, di-deprotonated and fully deprotonated forms.
- Additional tables showing relative free energies and proportions of all conformations of citrate in water for charge states -1 and -2 obtained by MP2 calculations.

This material is available free of charge via the Internet at <http://pubs.acs.org/>.

## References

- (1) Anastassiadis, S.; Morgunov, I.G.; Kamzolova, S.V.; Finogenova, T.V. Citric Acid Production Patent Review. *Recent Pat. Biotechnol.* **2008**, *2*, 107-123.
- (2) Pundir, R. K.; Jain, P. Evaluation of Five Chemical Food Preservatives for their Antibacterial Activity against Bacterial Isolates from Bakery Products and Mango Pickles. *J. Chem. Pharm. Res.* **2011**, *3*, 24-31.
- (3) Ling, C.; Liu, F.; Pei, Z.; Zhang, X.; Wei, M.; Zhang, Y.; Zheng, L.; Zhang, J.; Li, A.; Xing, B. Citric Acid Enhanced Copper Removal by a Novel Multi-amines Decorated Resin. *Sci. Rep.* **2015**, *5*, 9944.
- (4) Kimling, J.; Maier, M.; Okenve, B.; Kotaidis, V.; Ballot H.; Plech, A. Turkevich Method for Gold Nanoparticle Synthesis Revisited. *J. Phys. Chem. B* **2006**, *110*, 15700-15707.
- (5) Sasaki, S.; Kubota, N.; Doki, N. Adsorption Isotherms of Citric Acid Acting as a Growth-Suppressor onto the (100) and (111) Faces of Sodium Chloride Crystals in Supersaturated Aqueous Solution. *Chem. Eng. Technol.* **2006**, *29*, 247-250.
- (6) Bates, R. G.; Pinching, G. D. Resolution of the Dissociation Constants of Citric Acid at 0 to 50°, and Determination of Certain Related Thermodynamic Functions. *J. Am. Chem. Soc.* **1949**, *71*, 1274-1283.



- (7) Bénézeth, P.; Palmer, D. A.; Wesolowski, D. J. Dissociation Quotients for Citric Acid in Aqueous Sodium Chloride Media to 150°C. *J. Solution Chem.* **1997**, *26*, 63-84.
- (8) Crea, F.; De Stefano, C.; Millero, F. J.; Sharma, V. K. Dissociation Constants for Citric Acid in NaCl and KCl Solutions and their Mixtures at 25°C. *J. Solution Chem.* **2004**, *11*, 1349-1366.
- (9) Kochergina, L. A.; Vasilev, V. P.; Krutov, D. V.; Krutova, O. N. A Thermochemical Study of Acid-Base Interactions in Aqueous Solutions of Citric Acid. *Russian J. Phys. Chem.* **2007**, *81*, 182-186.
- (10) Borkovec, M.; Koper, G. J. M. A Cluster Expansion Method for the Complete Resolution of Microscopic Ionization Equilibria from NMR Titrations. *Anal. Chem.* **2000**, *72*, 3272-3279.
- (11) Borkovec, M.; Jönsson, B.; Koper, G. J. M. In *Surface and Colloid Science*; Matijevic, E., Ed.; Plenum Press: New York, 2001; Vol. 16. Ch. 2.
- (12) Szakács, Z.; Krasznai, M.; Noszál, B. Determination of Microscopic Acid-Base Parameters from NMR-pH Titrations. *Anal. Bioanal. Chem.* **2004**, *378*, 1428-1448.
- (13) Loewenstein, A.; Roberts, J. D. The Ionization of Citric Acid Studied by the Nuclear Magnetic Resonance Technique. *J. Am. Chem. Soc.* **1960**, *82*, 2705-2710.
- (14) Martin, R. B. A Complete ionization Scheme for Citric Acid. *J. Phys. Chem.* **1961**, *65*, 2053-2055.
- (15) Pearce, K. N.; Creamer, L. K. The Complete Ionization Scheme for Citric Acid. *Aust. J. Chem.* **1975**, *28*, 2409-2415.
- (16) Borkovec, M.; Hamacek, J.; Piguet, C. Statistical Mechanical Approach to Competitive Binding of Metal Ions to Multi-Center Receptors. *Dalton Trans.* **2004**, 4096-4105.
- (17) Riss-Johannessen, T.; Favera, N. D.; Todorova, T. K.; Huber, S. M.; Gagliardi, L.; Piguet, C. Understanding, Controlling and Programming Cooperativity in Self-Assembled Polynuclear Complexes in Solution. *Chem. Eur. J.* **2009**, *15*, 12702-12718.

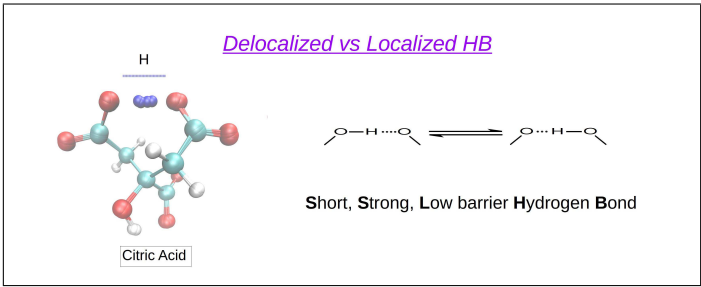
- (18) Noszal, B.; Sandor, P. Rota-Microspeciation of Aspartic Acid and Asparagine. *Anal. Chem.* **1989**, *61*, 2631-2637.
- (19) Takeda, Y.; Samejima, K.; Nagano, K.; Watanabe, M.; Sugeta H.; Kyogoku, Y. Determination of Protonation Sites in Thermospermine and in Some Other Polyamines by  $^{15}\text{N}$  and  $^{13}\text{C}$  Nuclear Magnetic Resonance Spectroscopy. *Eur. J. Biochem.* **1983**, *130*, 383-389.
- (20) Koper, G. J. M.; van Genderen, M. H. P.; Elissen-Román, C.; Baars, M. W. P. L.; Meijer, E. J.; Borkovec, M. Protonation Mechanism of Poly(propylene imine) Dendrimers and Some Associated Oligo Amines. *J. Am. Chem. Soc.* **1997**, *119*, 6512-6521.
- (21) Borkovec, M.; Spiess, B. Microscopic Ionization Mechanism of Inositol Tetrakisphosphates. *Phys. Chem. Chem. Phys.* **2004**, *6*, 1144-1151.
- (22) Riley, A. M.; Trusselle, M.; Kuad, P.; Borkovec, M.; Cho, J.; Choi, J. H.; Qian, X.; Shears, S. B.; Spiess, B.; Potter, B. V. L. *scyllo*-Inositol Pentakisphosphate as an Analogue of *myo*-Inositol 1,3,4,5,6-Pentakisphosphate: Chemical Synthesis, Physicochemistry and Biological Applications. *ChemBioChem* **2006**, *7*, 1114-1122.
- (23) Witanowski, M.; Stefaniak, L.; Webb, G. A. Nitrogen NMR Spectroscopy. *Annu. Rep. NMR Spectrosc.* **1987**, *18*, 1-211.
- (24) Moore, G. J.; Sillerud, L. O. The pH Dependence of Chemical Shift and Spin-Spin Coupling for Citrate. *J. Magnetic Resonance Series B* **1994**, *103*, 87-88.
- (25) Cleland, W. W. The Low-Barrier Hydrogen Bond in Enzymatic Catalysis. *Adv. Phys. Org. Chem.* **2010**, *44*, 1-17.
- (26) Ellison, R. D.; Levy, H. A. A Centered Hydrogen Bond in Potassium Hydrogen Chloromaleate: A Neutron Diffraction Structure Determination. *Acta Cryst.* **1965**, *19*, 260-268.
- (27) Steiner, T.; Majerz, I.; Wilson, C. C. First O-H-N Hydrogen Bond with a Centered Proton Obtained by Thermally Induced Proton Migration. *Angew. Chem.* **2001**, *40*, 2651-2654.

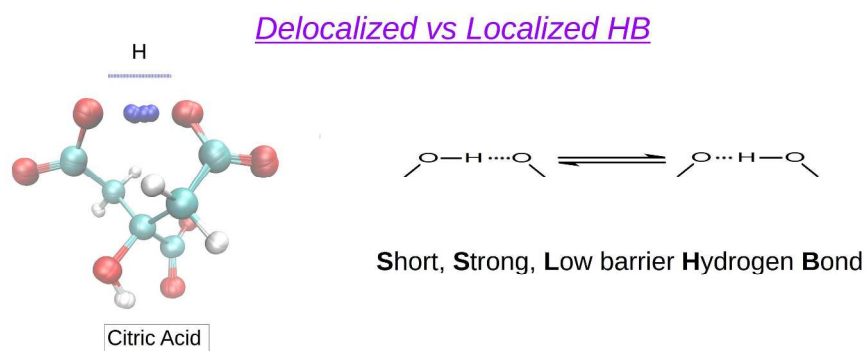
- (28) Perrin, C. L. Are Short, Low-Barrier Hydrogen Bonds Unusually Strong? *Acc. Chem. Res.* **2010**, *43*, 1550-1557.
- (29) Stasko, D.; Hoffmann, S. P.; Kim, K.-C.; Fackler, N. L. P.; Larsen, A. S.; Drovetskaya, T.; Tham, F. S.; Reed, C. A.; Richard, C. E. F.; Boyd, P. D. W.; Stoyanov, E. S. Molecular Structure of the Solvated Proton in Isolated Salts. Short, Strong, Low Barrier (SSLB) H-bonds. *J. Am. Chem. Soc.* **2002**, *124*, 13869-13876.
- (30) Reed, C. A. Myths about the Proton. The Nature of  $H^+$  in Condensed Media. *Acc. Chem. Res.* **2013**, *46*, 2567-2575.
- (31) Kawaguchi, S.; Kitano, T.; Ito, K. Dissociation Behaviour of Poly(fumaric acid) and Poly(maleic acid). 3. Infrared and Ultraviolet Spectroscopy. *Macromolecules* **1992**, *25*, 1294-1299.
- (32) Kolesnikov, S. P.; Lyudkovskaya, I. V.; Antipin, M. Yu.; Struchnov, T. Yu.; Neferov, O. M. Etherates of Friedel-Crafts Acids with a Short Hydrogen Bond: Symmetrical  $[Et_2O \cdots OEt_2]^+$  Cation in the Crystal Structure of the Etherate  $(Et_2O)_2 \cdot HZnCl_3$ . *Bull. Acad. Sci. USSR, Chem. Sci.* **1985**, *34*, 74-80.
- (33) Perrin, C. L. Symmetries of Hydrogen Bonds in Solution. *Science*, **1994**, *266*, 1665-1668.
- (34) Garcia-Viloca, M.; González-Lafont, A.; Lluch, J. M. Asymmetry of the Hydrogen Bond of Hydrogen Phthalate Anion in Solution. A QM/MM Study. *J. Am. Chem. Soc.* **1999**, *121*, 9198-9207.
- (35) Khan, M. A. S.; Sen, A.; Ganguly, B. Probing the Influence of pH Dependent Citric Acid Towards the Morphology of Rock Salt: a Computational Study. *CrystEngComm* **2009**, *11*, 2660-2667.
- (36) Wright, L. B.; Rodger, P. M.; Walsh, T. R. Aqueous Citrate: a First-Principles and Force-Field Molecular Dynamics Study. *RCS Adv.* **2013**, *3*, 16399-16409.

- (37) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378-6396.
- (38) Garcés, J. L.; Koper, G. J. M.; Borkovec, M. Ionization Equilibria and Conformational Transitions in Polyprotic Molecules and Polyelectrolytes. *J. Phys. Chem. B* **2006**, *110*, 10937-10950.
- (39) Garcés, J. L.; Madurga, S.; Borkovec, M. Coupling of Conformational and Ionization Equilibria in Linear Poly(ethylenimine): a Study Based on the Site Binding/Rotational Isomeric State (SBRIS) Model. *Phys. Chem. Chem. Phys.* **2014**, *16* 4626-4638.
- (40) Slichter, C. P. *Principles of Magnetic Resonance*. Springer. Berlin, 1980.
- (41) Flory, P. L. *Statistical Mechanics of Chain Molecules*; John Wiley: New York, 1969.
- (42) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Gaussian 09, revision D01, 2012; Gaussian, Inc.: Wallingford, CT, 2009.
- (43) Kolar, M.; Fanfrlik, J.; Lepsik, M.; Forti, F.; Luque, F. J.; Hobza, P. Assessing the Accuracy and Performance of Implicit Solvent Models for Drug Molecules: Conformational Ensemble Approaches. *J. Phys. Chem. B* **2013**, *117*, 5950-5962.
- (44) French, A. D.; Johnson, G. P.; Cramer, C. J.; Csonka, G. I. Conformational Analysis of Cellobiose by Electronic Structure Theories. *Carbohydrate Research* **2012**, *350*, 68-76.
- (45) Yao, G.; Zhang, J.; Huang, Q. Conformational and Vibrational Analyses of Meta-Tyrosine: an Experimental and Theoretical Study. *Spectrochim. Acta Mol. Biomol. Spectrosc.* **2015**, *151*, 111-123.
- (46) Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Use of Solution-Phase Vibrational Frequencies in Continuum Models for the Free Energy of Solvation. *J. Phys. Chem B* **2011**, *115*, 14556-14562.

- (47) MATLAB version 7.10.0. Natick, Massachusetts: The MathWorks Inc., 2010.
- (48) Alper, J. S.; Gelb, R. I. Standard Errors and Confidence Intervals in Nonlinear Regression: Comparison of Monte Carlo and Parametric Statistics. *J. Phys. Chem.* **1990**, *94*, 4747-4751.
- (49) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery B. P. *Numerical Recipes in Fortran 77: the art of scientific computing* 1986-1992. Cambridge University Press. Ch. 15.
- (50) De Robertis, A.; De Stefano, C.; Rigano, C.; Sammartano, S. Thermodynamic Parameters for the Protonation of Carboxylic Acids in Aqueous Tetraethylammonium Iodide Solutions. *J. Solution Chem.* **1990**, *19*, 569-587.
- (51) de Groot, J.; Koper, G. J. M.; Borkovec, M.; de Bleijser, J. Dissociation Behavior of Poly(maleic acid): Potentiometric Titrations, Viscometry, Pulsed Field Gradient NMR, and Model Calculations. *Macromolecules* **1998**, *31*, 4182-4188.
- (52) Smits, R. G.; Koper, G. J. M.; Mandel, M. The Influence of Nearest- and Next-Nearest-Neighbor Interactions on the Potentiometric Titration of Linear Poly(ethylenimine). *J. Phys. Chem.* **1993**, *97*, 5745-5751.
- (53) Lugo, M. L.; Lubes, V. R. Ternary Complex Formation between Chromium(III)-Picolinic Acid, Chromium(III)-Dipicolinic Acid, and Small Blood Serum Biologands. *J. Chem. Eng. Data* **2007**, *52*, 1217-1222.
- (54) Wyrzykowski, D.; Czupryniak, J.; Ossowski, T.; Chmurzynski, L. Thermodynamic Interactions of the Alkaline Earth Metal Ions with Citric Acid. *J. Therm. Anal. Calorim.* **2010**, *102*, 149-154.
- (55) Camoes, M. F.; Lito, M. J. G.; Ferra, M. I. A.; Covington, A. K. Consistency of pH Standard Values with the Corresponding Thermodynamic Acid Dissociation Constants. *Pure&Appl. Chem.* **1997**, *69*, 1325-1333.
- (56) Schreckenbach, G. Differential Solvation. *Chem. Eur. J.* **2017**, *23*, 3797-3803.

Graphical TOC Entry





Delocalized Proton Binding of Citric Acid

900x350mm (96 x 96 DPI)

## Supporting Information

### Ionization and Conformational Equilibria of Citric Acid: Delocalized Proton Binding in Solution

Sergio Madurga<sup>a\*</sup>, Miroslava Nedyalkova<sup>b</sup>, Francesc Mas<sup>a</sup> and Josep Lluís Garcés<sup>c</sup>

<sup>a</sup> Materials Science and Physical Chemistry Department & Research Institute of Theoretical and Computational Chemistry (IQTUB) of Barcelona University (UB), C/ Martí i Franquès, 1. 08028 Barcelona (Catalonia, Spain)

<sup>b</sup> Inorganic Chemistry Department, Faculty of Chemistry and Pharmacy “St Kliment Ohridski”, University of Sofia, 1 James Bourchier Blvd. 1164, Sofia (Bulgaria)

<sup>c</sup> Chemistry Department and AGROTECNIO, University of Lleida (UdL), Rovira Roure, 191, 25198, Lleida (Catalonia, Spain)

\* corresponding author. e-mail: s.madurga@ub.edu; [www.ub.edu/biophyschem](http://www.ub.edu/biophyschem)

**The reported best-fitted parameters correspond to the unique local minimum of the fitting procedure. An additional argument to the robustness of the found microscopic binding parameters.**

Let us outline an argument to show that the best-fitted parameters shown in this work correspond to a unique minimum of the fitting process. Firstly note that by rewriting the equations

$$\begin{aligned}\delta &= \delta_0 + \delta_{13}\theta_1 + (\delta_P - \delta_0 - \delta_{13})\theta_2 \\ \Delta &= \Delta_0 + \Delta_{13}\theta_1 + (\Delta_P - \Delta_0 - \Delta_{13})\theta_2\end{aligned}\tag{1}$$

replacing

$$\begin{cases} \theta_1(a_H) = \pi_1(H\Theta\Theta)P_1(a_H) + (1 - \pi_2(HH\Theta))P_2(a_H) + P_3(a_H) \\ \theta_2(a_H) = (1 - 2\pi_1(H\Theta\Theta))P_1(a_H) + 2\pi_2(HH\Theta)P_2(a_H) + P_3(a_H) \end{cases}\tag{2}$$

in eqn. (1),  $\delta$  and  $\Delta$  can be expressed in terms of the macroscopic probabilities  $P_i(a_H)$ , which only depend on the macroscopic dissociation  $pK_i$ -values *via*

$$P_n(a_H) = \frac{\overline{K}_n a_H^n}{\Xi}; \quad \Xi = \sum_{i=0, \dots, N} \overline{K}_i a_H^i; \quad pK_i = -\log K_i = \log(\overline{K}_{N-i+1}/\overline{K}_{N-i})\tag{3}$$

Introducing Eqns. (2) in (1) we obtain  $\delta$  and  $\Delta$  as a linear combination of the macroscopic probabilities. As a result we obtain the equations

$$\begin{aligned}\delta &= \delta_0 + X P_1(a_H) + Y P_2(a_H) + (\delta_P - \delta_0) P_3(a_H) \\ \Delta &= \Delta_0 + x P_1(a_H) + y P_2(a_H) + (\Delta_P - \Delta_0) P_3(a_H)\end{aligned}\tag{4}$$



where  $X$ ,  $Y$ ,  $x$  and  $y$  are new parameters to be fitted and defined as

$$\begin{aligned} X &= (2\delta_0 - 2\delta_P + 3\delta_{13}) \pi_1(\text{H}\Theta\Theta) + (\delta_P - \delta_0 - \delta_{13}) \\ Y &= (2\delta_P - 2\delta_0 - 3\delta_{13}) \pi_2(\text{H}\Theta\Theta) + \delta_{13} \\ x &= (2\Delta_0 - 2\Delta_P + 3\Delta_{13}) \pi_1(\text{H}\Theta\Theta) + (\Delta_P - \Delta_0 - \Delta_{13}) \\ y &= (2\Delta_P - 2\Delta_0 - 3\Delta_{13}) \pi_2(\text{H}\Theta\Theta) + \Delta_{13} \end{aligned} \quad (5)$$

But we have very good values for the macroscopic dissociation constants, obtained previously by potentiometry. Recalling too that  $\delta_P$ ,  $\delta_0$ ,  $\Delta_P$  and  $\Delta_0$  do not need to be fitted (they are directly obtained from the chemical shifts at very low and very high pH-values), we get that Eqns. (4) reduce to a *linear* fitting of  $X$ ,  $Y$ ,  $x$  and  $y$ . As a result, only *one minimum* can be obtained in fitting eqn. (4) to the experimental NMR curves. Thus, the values of  $X$ ,  $Y$ ,  $x$  and  $y$  are uniquely determined. The obtained values are (within the experimental and numerical error)  $X=0.137$ ,  $Y=0.266$ ,  $x=-15.14$ ,  $y=-5.65$ .

With those values for  $X$ ,  $Y$ ,  $x$  and  $y$  at hand, the problem reduces to show that the solution of the system of Eqns. (5), containing four unknowns ( $\pi_1(\text{H}\Theta\Theta)$ ,  $\pi_2(\text{H}\Theta\Theta)$ ,  $\delta_{13}$  and  $\Delta_{13}$ ) is unique for values of  $\pi_1(\text{H}\Theta\Theta)$  and  $\pi_2(\text{H}\Theta\Theta)$  lying between 0 and 1. This is an elementary exercise of algebra and numerical computation by eliminating  $\delta_{13}$  and  $\Delta_{13}$  from the system of equations and analysing the solutions for the resulting 2-equations system for  $\pi_1(\text{H}\Theta\Theta)$  and  $\pi_2(\text{H}\Theta\Theta)$ . It is found that the only solution (within the experimental and numerical error) corresponds to

$$\pi_1(\text{H}\Theta\Theta) \simeq 0.5; \pi_2(\text{H}\Theta\Theta) \simeq 0.20 \quad (6)$$

which are very close to the obtained values by direct fitting of all the microscopic binding parameters to the experimental NMR shift data. Furthermore, recalling that

$$\pi_2(\text{H}\Theta\Theta) = \frac{1}{2 + \frac{k_2}{k_1}} \simeq 0.20 \Rightarrow k_2/k_1 \simeq 3 \quad (7)$$

which is very close to the ratio obtained by direct fitting. Replacing (7) in the expression for  $\pi_1(\text{H}\Theta\Theta)$

$$\pi_1(\text{H}\Theta\Theta) = \frac{1}{2 + \frac{k_1 v}{k_2 u}} \simeq 0.5 \Rightarrow \frac{v}{3u} \simeq 0 \quad (8)$$

so that the repulsive interaction between terminal charged groups must be larger than the interaction terminal-central group.

## FIGURES of Supplementary Information

Free energy minimum conformations of citric acid obtained from *ab initio* optimizations in solvent for fully protonated (Fig. S1), mono-deprotonated (Fig. S2), di-deprotonated (Fig. S3) and fully deprotonated (Fig S4) forms. In Fig S2 and S3, it is also indicated the labelling to specify the state of each of the three carboxylic groups: “H” specifies a protonated carboxylic group and “ $\Theta$ ” specifies a deprotonated carboxylic group. The participation of a carboxylic group in a intramolecular hydrogen bond is indicated by a overbar over the corresponding “H” or “ $\Theta$ ” site.

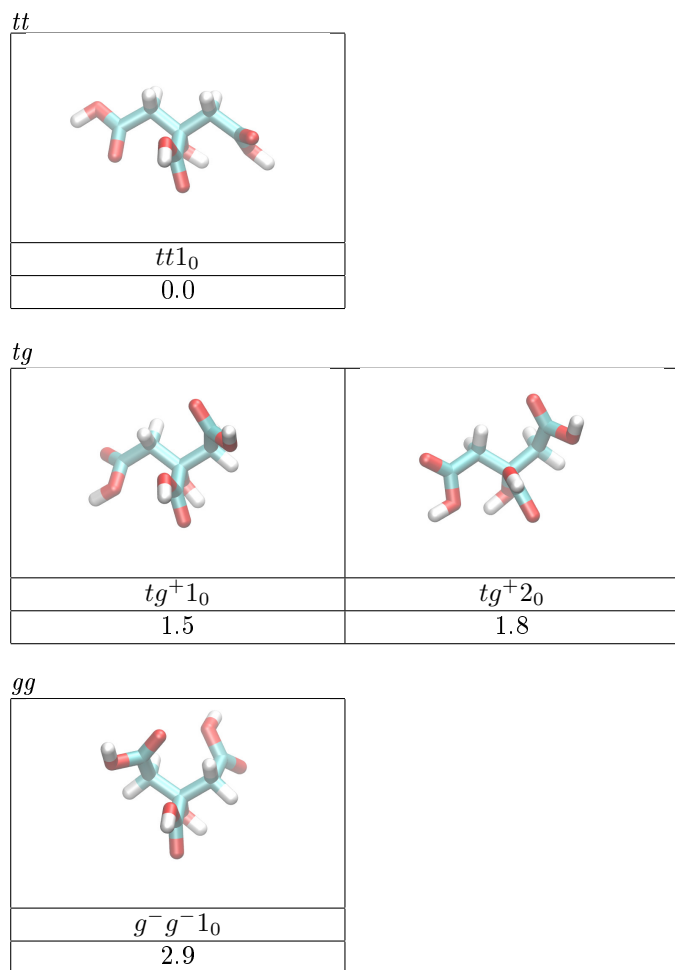


Figure S1: *tt*, *tg* and *gg* minimum conformations of citric acid obtained from optimizations at MP2 level with the SMD water model with the 6-311++G(d,p) basis set for charge state 0. The labeling indicates the conformation of each of the two principal dihedral angles (*t* for *trans*, *g* for *gauche*). The (+,-) superscript of *g* indicates the sign of the *gauche* angle. Relative free energies with respect to the most stable minimum conformation for each charge state in kcal/mol. The color of atoms correspond to blue for carbon, red for oxygen, white for hydrogen. Intermolecular hydrogen bonds are also indicated by red dashed lines.

<i>tt</i>		
<i>tt1</i> <sub>-1</sub> HΘH	<i>tt2</i> <sub>-1</sub> HΘH	<i>tt3</i> <sub>-1</sub> HΘH
0	1.5	1.6
<i>tt4</i> <sub>-1</sub> HΘH	<i>tt5</i> <sub>-1</sub> HΘH	<i>tt6</i> <sub>-1</sub> ΘHH
2.1	2.2	2.5
<i>tt7</i> <sub>-1</sub> ΘHH	<i>tt8</i> <sub>-1</sub> ΘHH	
2.6	2.7	

Figure S2: *tt* minimum conformations of citric acid obtained from optimizations at MP2 level with the SMD water model with the 6-311++G(d,p) basis set for charge state -1. The same preferences and units as indicated in Fig S1 were used.

*tg*

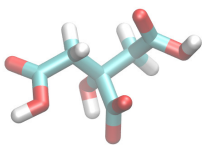
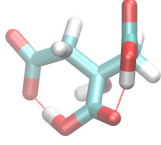
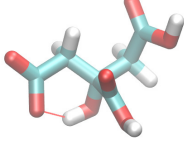
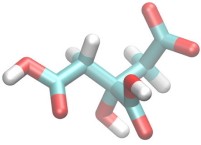
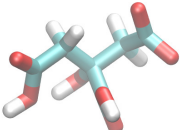
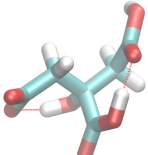
		
$tg^{+1-1} \text{H}\Theta\text{H}$	$tg^{+2-1} \overline{\Theta\text{H}\text{H}}$	$tg^{+3-1} \overline{\Theta\text{H}\text{H}}$
1.2	3.2	3.7
		
$tg^{+4-1} \text{H}\text{H}\Theta$	$tg^{+5-1} \text{H}\text{H}\Theta$	$tg^{+6-1} \overline{\Theta\text{H}\text{H}}$
4.9	6.8	7.0

Figure S3: *tg* minimum conformations of citric acid obtained from optimizations at MP2 level with the SMD water model with the 6-311++G(d,p) basis set for charge state -1. The same preferences and units as indicated in Fig S1 were used.

**Fig. S3.** Minimum conformations of citric acid obtained from optimizations at MP2 level with the SMD water model with the 6-311++G(d,p) basis set for charge state -2. The same preferences and units as indicated in Fig S1 were used.

*gg*

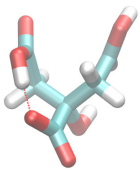
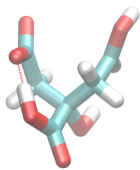
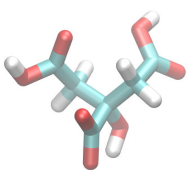
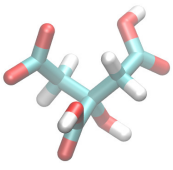
		
$g^-g^-1_{-1} \text{H}\Theta\text{H}$	$g^-g^-2_{-1} \Theta\text{HH}$	$g^-g^-3_{-1} \text{H}\Theta\text{H}$
0.0	0.9	2.6
		
$g^-g^-4_{-1} \Theta\text{HH}$		
8.4		

Figure S4: *gg* minimum conformations of citric acid obtained from optimizations at MP2 level with the SMD water model with th 6-311++G(d,p) basis set for charge state -1. The same preferences and units as indicated in Fig S1 were used.

*tt*

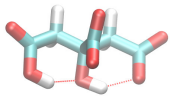
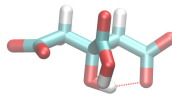
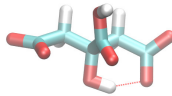
		
$tt1_{-2} \text{H}\Theta\Theta$	$tt2_{-2} \Theta\text{H}\Theta$	$tt3_{-2} \Theta\text{H}\Theta$
3.4	3.9	4.8

Figure S5: *tt* minimum conformations of citric acid obtained from optimizations at MP2 level with the SMD water model with th 6-311++G(d,p) basis set for charge state -2. The same preferences and units as indicated in Fig S1 were used.

*tg*

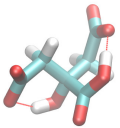
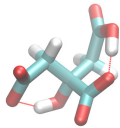
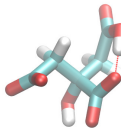
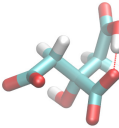
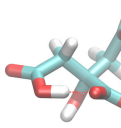
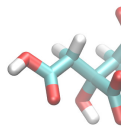
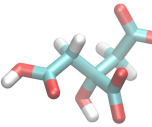
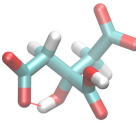
		
$tg^{+1-2} \Theta H \bar{\Theta}$	$tg^{+2-2} \bar{\Theta} \bar{\Theta} H$	$tg^{+3-2} \bar{\Theta} \bar{\Theta} H$
0.8	1.2	1.5
		
$tg^{+4-2} \bar{\Theta} \bar{\Theta} H$	$tg^{+5-2} H \bar{\Theta} \bar{\Theta}$	$tg^{+6-2} H \bar{\Theta} \bar{\Theta}$
3.5	5.0	5.0
		
$tg^{+7-2} H \bar{\Theta} \bar{\Theta}$	$tg^{+8-2} \bar{\Theta} H \bar{\Theta}$	
5.1	6.1	

Figure S6: *tg* minimum conformations of citric acid obtained from optimizations at MP2 level with the SMD water model with the 6-311++G(d,p) basis set for charge state -2. The same preferences and units as indicated in Fig S1 were used.

gg

$g^-g^-1_{-2} \text{H}\bar{\Theta}\bar{\Theta}$ 0.0	$g^-g^-2_{-2} \bar{\Theta}\bar{\Theta}H$ 0.3	$g^-g^-3_{-2} \bar{\Theta}H\bar{\Theta}$ 0.9
$g^-g^-4_{-2} \text{H}\bar{\Theta}\bar{\Theta}$ 1.0	$g^-g^-5_{-2} \text{H}\bar{\Theta}\bar{\Theta}$ 1.3	$g^-g^-6_{-2} \text{H}\bar{\Theta}\bar{\Theta}$ 2.8
$g^-g^-7_{-2} \text{H}\bar{\Theta}\bar{\Theta}$ 3.0	$g^-g^-8_{-2} \bar{\Theta}\bar{\Theta}H$ 3.3	$g^-g^-9_{-2} \bar{\Theta}\bar{\Theta}H$ 6.7
$g^-g^-10_{-2} \text{H}\bar{\Theta}\bar{\Theta}$ 7.6	$g^-g^-11_{-2} \bar{\Theta}H\bar{\Theta}$ 7.7	$g^-g^-12_{-2} \bar{\Theta}H\bar{\Theta}$ 11.1

Figure S7: *gg* minimum conformations of citric acid obtained from optimizations at MP2 level with the SMD water model with the 6-311++G(d,p) basis set for charge state -2. The same preferences and units as indicated in Fig S1 were used.



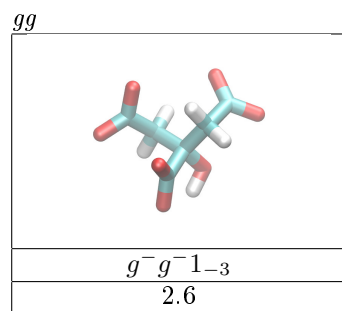
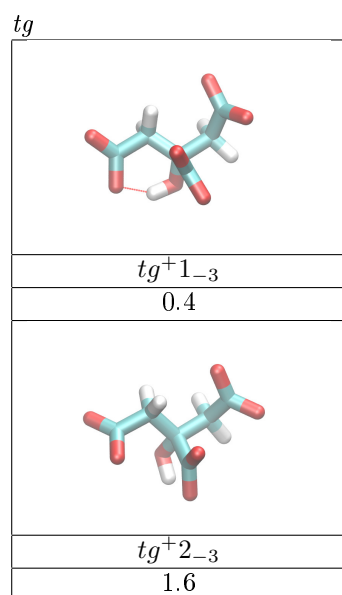
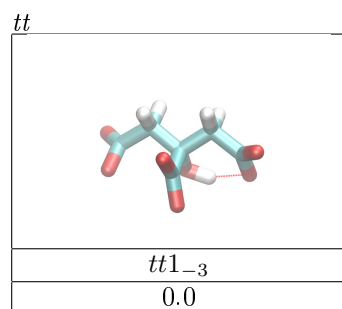


Figure S8: *tt*, *tg* and *gg* Minimum conformations of citric acid obtained from optimizations at MP2 level with the SMD water model with th 6-311++G(d,p) basis set for charge state -3. The same preferences and units as indicated in Fig S1 were used.

## TABLES of Supplementary Information

Tables S1 and S2 show relative free energies and proportions of all conformations of citrate in water for charge states -1 and -2 obtained by MP2 calculations.

Table S1: Relative Free energies (in kcal/mol) in water of different conformations for the mono-deprotonated form of citrate. Conformational proportions are calculated at 25°C. “H” specifies a protonated carboxylic group and “Θ” specifies a deprotonated carboxylic group. The participation of a carboxylic group in a intramolecular hydrogen bond between two carboxylic groups is indicated by an overbar over the corresponding “H” or “Θ” site.

conformation		$\Delta G_{rel}$	Relative population
$tt1_{-1}$	HΘH	0	23%
$tt2_{-1}$	$\overline{H\Theta H}$	1.5	3.7%
$tt3_{-1}$	$\overline{H\Theta H}$	1.6	3.2%
$tt4_{-1}$	$\overline{H\Theta H}$	2.1	1.5%
$tt5_{-1}$	HΘH	2.2	1.1%
$tt6_{-1}$	ΘHH	2.5	0.8%
$tt7_{-1}$	ΘHH	2.6	0.6%
$tt8_{-1}$	ΘHH	2.7	0.5%
$tg^+1_{-1}$	HΘH	1.2	6.1%
$tg^+2_{-1}$	$\overline{\Theta HH}$	3.2	0.2%
$tg^+3_{-1}$	ΘHH	3.7	0.1%
$tg^+4_{-1}$	HHΘ	4.9	0.0%
$tg^+5_{-1}$	HHΘ	6.8	0.0%
$tg^+6_{-1}$	$\overline{\Theta HH}$	7.0	0.0%
$g^-g^-1_{-1}$	$\overline{H\Theta H}$	0.0	48%
$g^-g^-2_{-1}$	$\overline{\Theta HH}$	0.9	10%
$g^-g^-3_{-1}$	HΘH	2.6	0.6%
$g^-g^-4_{-1}$	ΘHH	8.4	0.0%

Table S2: Relative Free energies (in kcal/mol) in water of different conformations for the di-deprotonated form of citrate. Conformational proportions are calculated at 25°C. Charged carboxylic groups and intramolecular hydrogen bonds are indicated as in Table S1.

conformation		$\Delta G_{rel}$	Relative population
$tt1_{-2}$	H $\Theta\Theta$	3.4	0.1%
$tt2_{-2}$	$\Theta H\Theta$	3.9	0.1%
$tt3_{-2}$	$\Theta H\bar{\Theta}$	4.8	0.0%
$tg^+1_{-2}$	$\Theta\bar{H}\bar{\Theta}$	0.8	10%
$tg^+2_{-2}$	$\Theta\Theta\bar{H}$	1.2	5.3%
$tg^+3_{-2}$	$\Theta\Theta\bar{H}$	1.5	3.1%
$tg^+4_{-2}$	$\Theta\Theta\bar{H}$	3.5	0.1%
$tg^+5_{-2}$	$\bar{H}\Theta\Theta$	5.0	0.0%
$tg^+6_{-2}$	H $\Theta\Theta$	5.0	0.0%
$tg^+7_{-2}$	H $\Theta\Theta$	5.1	0.0%
$tg^+8_{-2}$	$\Theta H\Theta$	6.1	0.0%
$g^-g^-1_{-2}$	$\bar{H}\Theta\bar{\Theta}$	0.0	37%
$g^-g^-2_{-2}$	$\bar{\Theta}\Theta\bar{H}$	0.3	24%
$g^-g^-3_{-2}$	$\bar{\Theta}\bar{H}\Theta$	0.9	8.3%
$g^-g^-4_{-2}$	$\bar{H}\bar{\Theta}\Theta$	1.0	6.8%
$g^-g^-5_{-2}$	$\bar{H}\bar{\Theta}\Theta$	1.3	4.0%
$g^-g^-6_{-2}$	$\bar{H}\bar{\Theta}\Theta$	2.8	0.3%
$g^-g^-7_{-2}$	$\bar{H}\Theta\bar{\Theta}$	3.0	0.3%
$g^-g^-8_{-2}$	$\bar{\Theta}\Theta\bar{H}$	3.3	0.1%
$g^-g^-9_{-2}$	$\Theta\Theta H$	6.7	0.0%
$g^-g^-10_{-2}$	H $\Theta\Theta$	7.6	0.0%
$g^-g^-11_{-2}$	$\Theta H\Theta$	7.7	0.0%
$g^-g^-12_{-2}$	$\Theta H\Theta$	11.1	0.0%